A formal approach to attitude

Dalege, J.

Publication date
2020

Document Version
Final published version

License
Other

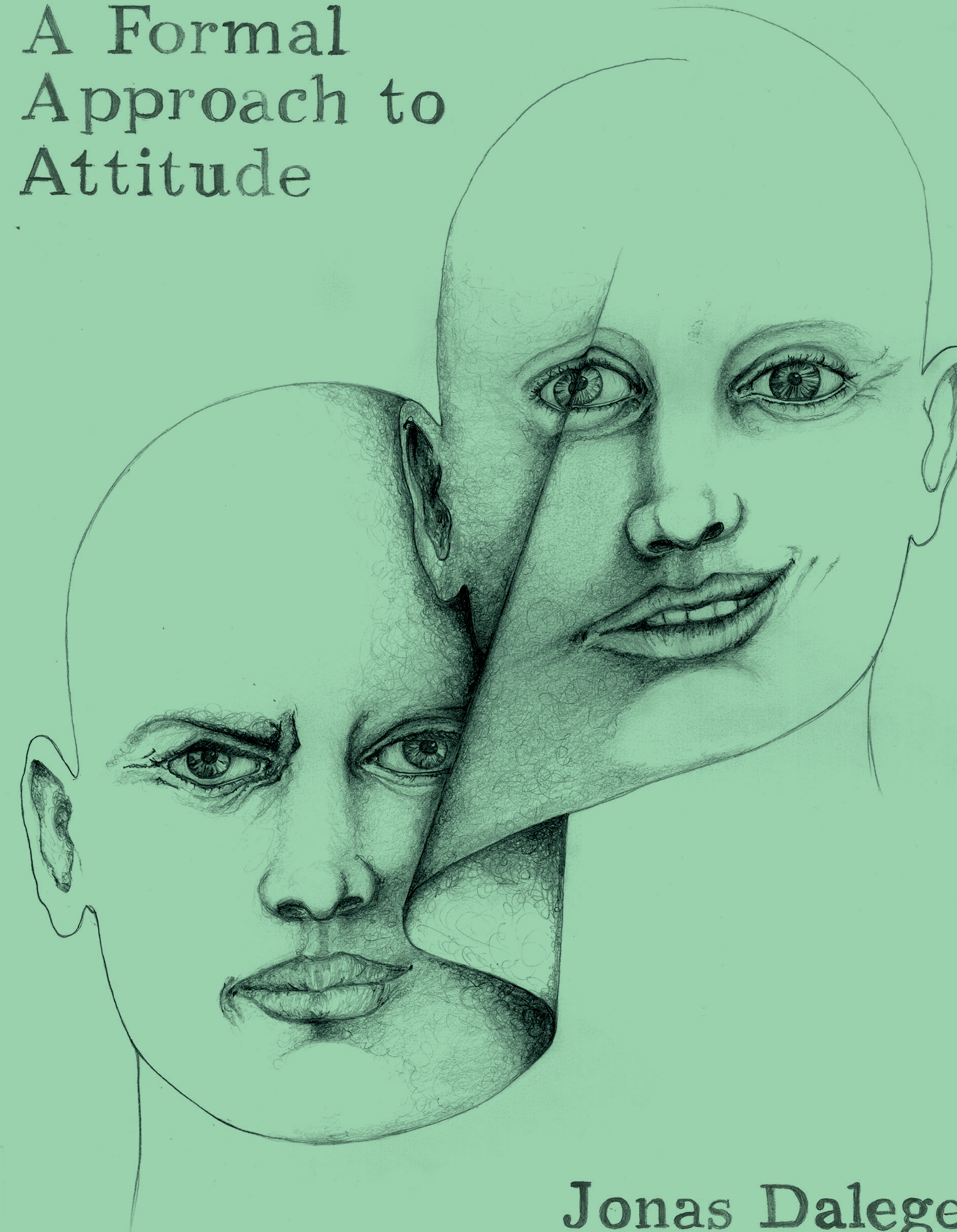
A Formal Approach to Attitude

Jonas Dalege

A Formal Approach to Attitude

Jonas Dalege

A Formal
Approach to
Attitude



Jonas Dalege

jdalege.com

Cover design: Andrea Frahn

This thesis was typeset using LaTeX and the thesis style design is based on the style design developed by Martijn Wieling (www.martijnwieling.nl).

A Formal Approach to Attitude

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. K.I.J. Maex

ten overstaan van een door het College voor Promoties ingestelde commissie,

in het openbaar te verdedigen in de Agnietenkapel

op donderdag 23 januari 2020, te 14:00 uur

door Jonas Dalege

geboren te Duisburg

Promotiecommissie

Promotor:	Prof. dr. Han L. J. van der Maas	Universiteit van Amsterdam
Copromotores:	Prof. dr. Denny Borsboom	Universiteit van Amsterdam
	Prof. dr. Frenk van Harreveld	Universiteit van Amsterdam
Overige leden:	Dr. Mark J. Brandt	Tilburg University
	Prof. dr. Juliane Degner	Universität Hamburg
	Prof. dr. Agneta H. Fischer	Universiteit van Amsterdam
	Prof. dr. Gerben A. van Kleef	Universiteit van Amsterdam
	Dr. Lourens J. Waldorp	Universiteit van Amsterdam
Faculteit:	Faculteit der Maatschappij- Gedragswetenschappen	

Contents

1 General Introduction **1**
1.1 Introduction 1
1.2 Outline 6

I The Causal Attitude Network (CAN) Model as a Measurement Model of Attitude **9**

2 The Causal Attitude Network (CAN) Model **11**
2.1 Introduction 12
2.1.1 The Need for a Formalized Measurement Model of Attitudes 14
2.1.2 From Cognitive Consistency to Network Models . . 17
2.2 The Causal Attitude Network (CAN) Model 20
2.2.1 Attitude Formation and Structure 21
2.2.2 Attitude Stability and Change 27
2.2.3 Attitude Strength 33
2.3 Discussion 40
2.3.1 Extensions of the CAN Model 41
2.3.2 The Challenge to Assess Attitude Networks 42
2.3.3 Future Study of the CAN Model 44
2.3.4 Conclusion 48

3 Network Analysis on Attitudes **49**
3.1 Introduction 50
3.1.1 What is a Network? 50
3.1.2 Network Analysis 52
3.1.3 Estimation of Attitude Networks 53
3.1.4 Community Detection 55
3.1.5 Node Centrality 57
3.1.6 Network Connectivity 58
3.1.7 Comparison of Networks 60

3.2 Network Simulation 61
3.2.1 Network Connectivity 62
3.2.2 Node Centrality 63
3.2.3 Testing Predictions from Simulations 65
3.3 Discussion 65

II Empirical Tests of the Causal Attitude Network (CAN) Model 67

4 A Network Perspective on Attitude Strength: Testing the Connectivity Hypothesis 69
4.1 Introduction 70
4.2 The Causal Attitude Network (CAN) Model 70
4.3 Global Connectivity of Attitude Networks 71
4.4 Network Connectivity and Attitude Strength 72
4.5 Overview of the Present Research 74
4.6 Method 74
4.6.1 Participants 74
4.6.2 Measures 75
4.6.3 Data Analysis 78
4.7 Results 79
4.8 Discussion 80
4.8.1 Limitations 82
4.8.2 Implications and Directions for Future Research . . 83
4.8.3 Conclusion 85

5 Network Structure Explains the Impact of Attitudes 87
5.1 Introduction 88
5.2 Results 93
5.2.1 Simulation Study 93
5.2.2 Test of Connectivity Hypothesis 94
5.2.3 Test of Centrality Hypothesis 95
5.2.4 Forecast Analysis 98
5.3 Discussion 98
5.4 Methods 101
5.4.1 Simulation of Networks 101
5.4.2 Participants 102
5.4.3 Measures 102
5.4.4 Statistical analyses 103

III The Attitudinal Entropy (AE) Framework as a General Theory of Attitude **107**

6 The Attitudinal Entropy (AE) Framework **109**
6.1 Introduction 110
6.2 The Attitudinal Entropy (AE) Framework 111
6.2.1 Attitudinal Micro- and Macrostates 111
6.2.2 Attitudinal Entropy 112
6.2.3 The Causal Attitude (CAN) Model 115
6.3 Levels of Attitudinal Entropy Reduction 120
6.3.1 Implicit Measures are more Likely to Tap Attitudes in High Entropy States 122
6.3.2 The Mere Thought Effect as an Initial Level of Heightened Attitudinal Entropy Reduction 124
6.3.3 Attitude Strength 126
6.3.4 Heuristic-Based vs. Argument-Based Persuasion as Global vs. Specific Threshold Changes 129
6.4 Aversiveness of Attitudinal Entropy: Ambivalence and Cognitive Dissonance from the Perspective of the AE Framework 131
6.4.1 Boltzmann Entropy as Ambivalence 132
6.4.2 Gibbs Entropy as Ambivalence 134
6.4.3 Cognitive Dissonance and Ambivalence Reflect Attitudinal Entropy 137
6.5 Future Study of the Attitudinal Entropy (AE) Framework . 138
6.5.1 Possible Neural Substrates of the AE Framework . . 138
6.5.2 The AE Framework's Relation to Other Models of Attitude 139
6.5.3 Open Questions 140
6.6 Conclusion 142

7 The Attitudinal Entropy (AE) Framework: Clarifications, Extensions, and Future Directions **143**
7.1 Prelude 144
7.2 Commentaries 144
7.3 Introduction 151
7.4 The Aim of the Attitudinal Entropy (AE) Framework 152
7.5 Principles and Definitions of the AE Framework 154
7.6 Beyond Dual-Process Models 158
7.6.1 Simulation R1: Modeling Impact of Argument Quantity Versus Argument Quality 159

7.7 Does the AE Framework Provide Novel Predictions? 162
7.8 Let's go All Psychology on Attitude Measurement 164
7.9 Extending the AE Framework to the Interpersonal Realm . 166
7.10 Attitudinal Entropy Framework – The App 168
7.11 An Essential Next Step 169

8 The Learning Ising Model of Attitude (LIMA): Entropy Reduction by Hebbian Learning 171
8.1 Introduction 172
8.2 The Learning Ising Model of Attitude (LIMA) 177
8.2.1 State Updating 177
8.2.2 Edge Weights Updating 178
8.3 Simulations 178
8.3.1 Simulation 1: Development of an Initial Attitude Network 179
8.3.2 Simulation 2: Decrease of Entropy through Learning 180
8.3.3 Simulation 3: Coupling of Instability and Attention 181
8.3.4 Simulation 4: Learning of Thresholds 184
8.3.5 Simulation 5: Attitude Change 186
8.4 Discussion 187

IV Conlusion 193

9 General Discussion 195
9.1 Introduction 195
9.2 CAN Model, AE Framework, and LIMA – What Is Needed to Explain What? 196
9.3 Limitations and Extensions 197
9.3.1 Intra- and Interpersonal Extensions 198
9.3.2 Temporal Dynamics 199
9.3.3 Methodological Developments 200
9.4 The Nature of Attitude Revisited 201
9.4.1 Measurement 202
9.4.2 Mental Causation 205
9.4.3 Structural Realism 207
9.5 Conclusion 208

V Appendices **209**

References **211**

A Description of the ANES 1984 Data Set **243**
A.1 Evaluative Reactions 243
A.2 Missing Data 243

B Supplementary Information on Chapter 3 **245**
B.1 Estimating Networks from Continuous or Categorical Data 245
B.2 Inclusion of Covariates 246
B.3 Edge Stability 247
B.4 Illustration of Shortest Path Lengths 248
B.5 Centrality Stability 249

C Supplementary R Code to Chapter 3 **253**

D Supplementary Notes on Chapter 5 **261**
D.1 Supplementary Note 1: Analytical Solutions of the Hypotheses 261
D.1.1 Logistic Regression and the Ising Model 261
D.1.2 Monotonicity of Prediction Loss as a Function of Connectivity 262
D.1.3 Closeness and Correlations 264
D.2 Supplementary Note 2: Alternative Analysis on Non-Voters 265
D.3 Supplementary Note 3: Alternative Analysis on Independents 265
D.4 Supplementary Note 4: Alternative Analysis on Missing Values 267
D.5 Supplementary Note 5: Alternative Analysis on Networks Based on Different Numbers of Attitude Elements 268
D.6 Supplementary Note 6: Alternative Analysis on Latent Variable Models 270

E Supplementary Information on Simulation 1b of Chapter 6 **275**

F Contributions to Chapters **277**

G Summary **281**
G.1 A Formal Approach to Attitude 281

H Nederlandse Samenvatting **285**
H.1 Een Formele Benadering van Attitude 285

I Acknowledgments – Dankwoord 289

Chapter 1

General Introduction

1.1 Introduction

> I can calculate the motion of heavenly bodies but not the madness of men. – Attributed to Sir Isaac Newton

There is little doubt that attitudes are immensely complex. Attitudes are formed through several different processes of different nature (e.g., Fazio, 1995, 2007; Zanna & Rempel, 1988), they can be highly stable and impactful or fluctuating and inconsequential (e.g., Krosnick & Petty, 1995), and are at the core of several central concepts within psychology. For example, prejudice can be defined as one's attitude toward an out-group, love is basically an extremely positive attitude toward another person, and self-esteem reflects one's attitude toward oneself. Furthermore, one would be hard-pressed to come up with any topic that has received more attention from research in the social sciences as attitudes have. For example, Google Scholar returns 3,900,000 hits for the search term "attitude" as of December, 2018. However, a general and formal theory – a theory that explains the phenomena of its subject matter in unambiguous mathematical or algorithmic terms – of attitude is yet to be found.

There is, of course, reason to think that it is impossible to capture the complexity of attitude in a rigorous formal fashion, which is also reflected in the standard view on the sciences that the further one moves away from the natural sciences, the less likely it is that a formal theory can be successful. It is undeniable that physics has been the most successful science when it comes to explaining its subject matter in a formal fashion, illustrated by the fact that a single equation – Schrödinger's equation – provides a comprehensive description of the (sub-)atomic level and a set of equations – Einstein's field equations of his general relativity theory – explains basically everything else. Moving further up the hierarchy of the sciences[1] to chemistry, we still find laws, such as the conversation of

[1]This hierarchy is based on the work by the philosopher Auguste Comte (1842), who

mass, but formal theories with such explanatory breadth as in physics are elusive. Similarly, biology's grand theory of evolution by natural selection still provides laws, but in a less formal way. Finally, arriving at psychology we generally find neither laws (except perhaps at very basic levels, such as perception) nor truly formal theories and some have even argued that formalization of psychological theories can hinder theoretical progress, because psychological science is too complex and immature (Higgins, 2004; Fiske, 2004; Lewin, 1951; see also Smaldino, 2017). Especially in the subfields of psychology that Paul E. Meehl (1978) called "soft psychology" (e.g., clinical and social psychology), efforts to develop formal theories are mostly the exception that proves the rule and it might indeed be the case that it is impossible to calculate the madness of men. However, another possibility is that up until now, the right lens through which to look at psychological processes to allow the development of formal theories has remained elusive. The lens that I use in this dissertation to develop and test several formalisms on one of psychology's core constructs – attitude – is that of analogical modeling (e.g., Haig, 2005). By applying concepts from statistical mechanics, I propose a formal approach to attitude relying on few first principles.

While there are valid reasons to think that psychology is just too complex to be captured in a parsimonious and formal framework, my view is that there is also ample evidence to the contrary: In many ways, humans are strikingly predictable and not that mad at all. Surely, humans can behave in highly unpredictable and erratic ways, but much of human life follows relatively predictable patterns. Whether one judges humans to be unpredictable again depends on the lens one uses – trying to predict what specific thought a person has at a given time is and will in all likelihood remain a hopeless endeavor, but predicting, for example, where a person will go at a specific point in time can be achieved with relatively high accuracy (e.g., Song, Qu, Blumm, & Barabsi, 2010). Advances in the formalization of psychological theories will therefore not only depend on explaining patterns but also on explaining what we cannot explain (e.g., explaining why in some instances attitude change is unpredictable). Just as with weather forecasts that are highly accurate in specifying their accuracy (e.g., it rains on roughly one in ten days when a weather forecast predicts that there is a 10% chance of rain), psychological theories should

argued that at the lowest level of the hierarchy are the physical sciences, because these sciences are the most simple and in these sciences it is the easiest to achieve consensus on general theories and laws. On the other end of the hierarchy are the social sciences, which are the most complex. That psychology can be placed above biology, chemistry, and physics in such a hierarchy also received empirical support (Simonton, 2004).

specify how likely it is their predictions will pan out.

A relatively recent development that provides reason for optimism about the possibility of developing formal theories in soft psychology is the growing field of complexity science (M. Mitchell, 2009). While complexity science consists of several different subfields, one of its central ideas is that simple rules at a lower level can lead to complex behavior at a higher level. As an illustration, take the flocking behavior of birds. Flocks of birds can show extremely complex behavior and looking at a flock of birds one almost cannot escape the thought that something is guiding this complex behavior. Yet, extremely simple models are able to reproduce flocking behavior by assigning three simple rules to the modeled birds' behavior (Reynolds, 1987). First, birds tend to align, implying that they are likely to fly in the same direction as their neighboring birds. Second, birds will turn away from a neighboring bird if they get too close. Third, unless they are not too close birds will move closer to nearby birds. Strikingly, these simple rules are able to reproduce flocking behavior, and illustrate that seemingly complex behavior at a higher level can emerge from simple rules at a lower level. Another point that this example illustrates is that, to understand the behavior of complex systems, one should not study the properties of the lower level entities in isolation (e.g. focusing on the molecular composition of wings); instead, one should focus on the interactions between lower level properties (i.e., the way the birds respond to each other's behavior). One could know basically everything there is to know about a single bird, but only by studying the consequences of the interactions between several birds it is possible to explain flocking behavior. Such phenomena were what led physicist Phillip Anderson to rightfully claim that "more is different" (P. W. Anderson, 1972, p. 393).

There has also been an increase in recent years to model complex phenomena in the soft psychology literature using relatively simple assumptions. For example, using an approach similar to the one taken in this dissertation, the mutualism model of intelligence is able to explain several hallmark findings in the intelligence literature, such as the hierarchical structure of intelligence (Carroll, 1993; Mackintosh, 1998) and the increasing heritability coefficient of intelligence with age (M. Bartels, Rietveld, van Baal, & Boomsma, 2002; Fulker, DeFries, & Plomin, 1988), based on the simple idea that different components of intelligence reinforce each other (e.g., reading skills lead to better understanding of mathematical rules; van der Maas et al., 2006). Similarly, assuming that symptoms of depression causally influence each other provides an explanation for the development and maintenance of depressive episodes (Cramer et al., 2016). Closer to the topic of the current dissertation, a recent model of social judgment provides a simple explanation of complex and seemingly contradictory findings in the social judgment literature (Galesic, Olsson, & Rieskamp, 2018). By assuming that individuals sample instances from memory to come up with social judgments (e.g., how many people hold a given political opinion) and that this sampling process interacts with properties in the environment, this Social Sampling Model (SSM) is able to account for false consensus (i.e., overestimating how many people share your belief) as well as for false uniqueness (i.e., underestimating how many people share your belief) and for instances of self-enhancement (i.e., overestimating your performance) as well as self-depreciation (i.e., underestimating your performance). Another impressive example of the ability to model complex psychological phenomena using rigorous and parsimonious formalizations is the application of Quantum Probability (QP) theory to judgment processes (Pothos & Busemeyer, 2013). While biases, such as the conjunction fallacy (i.e., judging a conjunction of two statements as more probable than one of the statements) or question-order effects (i.e., judging two questions differently depending on which question was asked first) are generally studied and explained in isolation from each other, QP theory is able to explain these different findings in a unified and parsimonious way. While still not being part of mainstream psychological science, the growing number of such theories illustrates that (soft) psychology is moving toward more formalized specifications of its core phenomena and the current dissertation aims to contribute to this development.

Before providing an overview of the approach to attitude articulated in this dissertation, I want to provide a brief discussion of the historically most prominent theories and models on attitude. Probably the first highly influential theory on attitude is the tripartite model of attitude (Rosenberg, Hovland, McGuire, & Brehm, 1960), in which an attitude (e.g., a negative evaluative disposition toward snakes) is treated as a latent entity that can express itself in cognitive (e.g., believing that snakes are dangerous), affective (e.g., being scared of snakes), and behavioral reactions (e.g., running away when one sees a snake). While the idea that a latent attitude causes these expressions has fallen out of vogue, there is still general agreement that attitudes consist of cognitive, affective, and behavioral components. For example, a more recent influential account holds that attitudes are object-evaluation associations in memory between an attitude object and a global evaluation that is based on cognitive, affective, and behavioral information (Fazio, 1995, 2007). An important aspect of this model is that the association between an attitude object and the global evaluation can vary in strength. If a global evaluation is strongly linked to the attitude object, this evaluation is activated automatically when one encounters the attitude object. Such attitudes can then be seen as stable representations in memory. However, whether attitudes are indeed stored in memory is still a matter of debate and other influential accounts assume that attitudes are temporary evaluations that are constructed on the spot (Schwarz, 2007). From this perspective, stable evaluations do not reflect a stable representation in memory but construction processes that may lead to similar results because of similar circumstances.

Related to the issue whether attitudes are global evaluations in memory and are therefore likely to be activated automatically (Fazio & Williams, 1986) is the debate in the attitude literature concerning the difference between implicitly versus explicitly measured attitudes (e.g., Fazio & Olson, 2003b; Gawronski, LeBel, & Peters, 2007; De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009). Typically, implicit measures of attitudes rely on reaction-time task, such as the popular Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), and thereby aim to assess the automatic activation of attitudes in an indirect way. Explicit or self-report measures, in contrast, directly ask participants about their attitude. Models on implicitly versus explicitly measured attitudes generally fall into two camps. First, based on the attitudes as object-evaluation association account, implicit measures are assumed to reflect a 'purer' measure of the attitude and explicit measures are assumed to reflect the outcome of the attitude and extra-attitudinal processes (e.g., social desirability concerns; Fazio & Olson, 2003b). Second, the relatively recent Associative-Propositional Evaluation (APE) model represents a dual-process account of attitudes, in which outcomes on implicit attitude measures represent associative 'gut-feelings' and outcomes on explicit measures represent propositional evaluations that are largely determined by consistency-pressures (e.g., one suppresses one's negative association of Blacks with danger, because one has the belief that prejudice is wrong; Gawronski & Bodenhausen, 2006).

While the different approaches to attitudes start from quite different assumptions, such as whether attitudes are stable representations in memory or constructed on the spot or whether a single process or two processes underlie evaluations, the empirical implications of these theories are often too flexible to test them against each other. For example, it is often unclear which empirical predictions follow from dual-process theories that do not follow from single-process theories (e.g., Corneille & Stahl, 2018; Ehret, Monroe, & Read, 2015; C. J. Mitchell, De Houwer, & Lovibond, 2009) or which predictions follow from assuming that attitudes are stored in memory that do not follow from assuming that attitudes are constructed on the spot (Schwarz, 2007). One aim of this dissertation therefore is to provide formalisms that make unambiguous empirical predictions and to provide a way to integrate these seemingly different views on attitudes.

The approach formulated in this dissertation starts with a simple assumption. Beliefs, feelings, and behaviors vis-á-vis an attitude object (i.e., attitude elements or evaluative reactions) not only provide the basis of an attitude but also directly interact with each other (e.g., judging snakes as dangerous leads to feeling afraid of snakes). These interactions can be modeled by representing them in a network structure and the global evaluation emerges from these interactions – in fact, only through these interactions a stable and unambiguous evaluation becomes possible, as I will show in Chapter 6. The second fundamental assumption of the approach is that the interactions between attitude elements become more pronounced as more attention and thought is directed at the attitude object, implying that attitudes are unstable and inconsistent (high in entropy) when one directs little attention to the attitude object, while attitudes become more stable and consistent (low in entropy) when one thinks frequently about the attitude object. Starting from these two assumptions, I derive formalisms on attitude that (a) provide a straightforward measurement model of attitude, (b) can explain a host of hallmark findings in the attitude literature, and (c) provide precise and unambiguous predictions.

1.2 Outline

The outline of this dissertation is as follows (see Figure 1.1 for an overview). In Part I, I introduce the Causal Attitude Network (CAN) model, which represents a measurement model of attitude that is based on the assumption that attitude elements form a network structure of causal interactions. Chapter 2 introduces the CAN model and discusses its implications for attitude formation and structure, attitude change and stability, and attitude strength. Chapter 3 provides a hands-on tutorial on how to apply the CAN model and network analysis to typical data on attitudes. Part II consists of two chapters providing empirical tests of some of the CAN model's predictions. In Chapter 4, I test the key prediction of the CAN model that highly connected attitude networks correspond to strong attitudes and show that indeed attitudes that are based on highly connected attitudes are stable and have high impact on behavior. In Chapter 5, I show that network structure of attitudes can be used to forecast how much impact an attitude has on behavior globally and which attitude elements have the highest impact specifically. Part III is devoted to extending the CAN model to a general theory of attitude – the Attitudinal Entropy (AE) framework, which assumes that interactions between attitude elements increase as more attention and thought is directed at the attitude object. Chapter 6 introduces the AE framework and shows that it can explain several hallmark findings in the attitude literature, such as the mere thought effect (Tesser, 1978) and effects of heuristic- versus argument-based persuasion (e.g., Chaiken, Liberman, & Eagly, 1989; Petty & Cacioppo, 1986). Chapter 7 provides a reply to several commentaries that were written in response to Chapter 6 and provides clarifications of the AE framework's assumptions, discusses dual-process models and attitude measurement from the perspective of the AE framework, and extends the AE framework to the interpersonal realm. Chapter 8 adds a learning mechanism to the AE framework and thereby provides an explanation for the development of attitude networks.

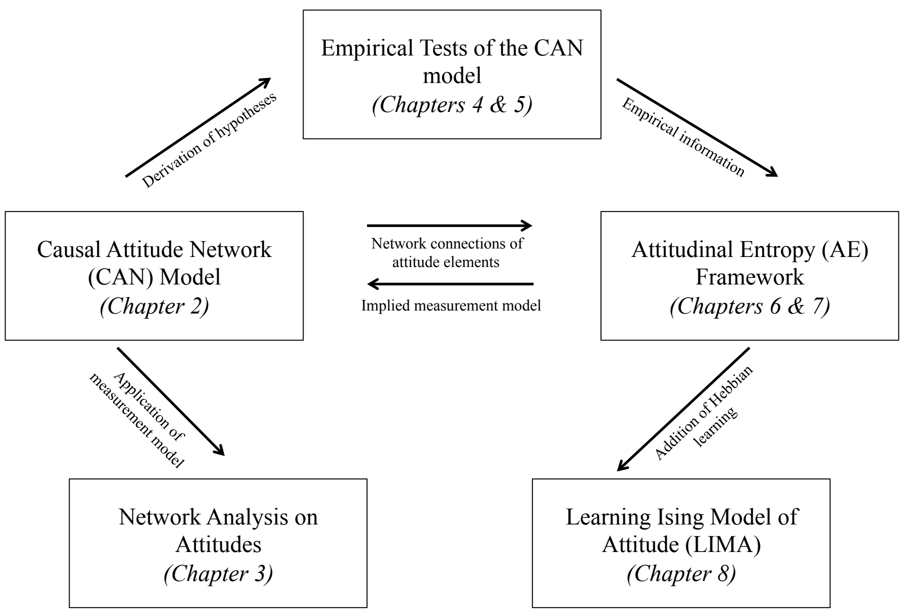


Figure 1.1: Overview of the dissertation.

Part I

The Causal Attitude Network (CAN) Model as a Measurement Model of Attitude

Chapter 2

The Causal Attitude Network (CAN) Model

Abstract

This chapter introduces the Causal Attitude Network (CAN) model, which conceptualizes attitudes as networks consisting of evaluative reactions and interactions between these reactions. Relevant evaluative reactions include beliefs, feelings, and behaviors toward the attitude object. Interactions between these reactions arise through direct causal influences (e.g., the belief that snakes are dangerous causes fear of snakes) and mechanisms that support evaluative consistency between related contents of evaluative reactions (e.g., people tend to align their belief that snakes are useful with their belief that snakes help maintain ecological balance). In the CAN model, the structure of attitude networks conforms to a small-world structure: evaluative reactions that are similar to each other form tight clusters, which are connected by a sparser set of 'shortcuts' between them. We argue that the CAN model provides a realistic formalized measurement model of attitudes and therefore fills a crucial gap in the attitude literature. Furthermore, the CAN model provides testable predictions for the structure of attitudes and how they develop, remain stable, and change over time. Attitude strength is conceptualized in terms of the connectivity of attitude networks and we show that this provides a parsimonious account of the differences between strong and weak attitudes. We discuss the CAN model in relation to possible extensions, implication for the assessment of attitudes, and possibilities for further study.

This chapter has been adapted from: Dalege, J., Borsboom, D., van Harreveld, F., van der Berg, H., Conner, M., & van der Maas, H. L. J. (2016). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychological Review, 123*, 2-22.

2.1 Introduction

The attitude concept continues to occupy a central role in the social sciences. Not only are attitudes a core topic in social psychology; they also play an important role in economics and the political sciences (e.g., Latané & Nowak, 1994). While research on attitudes has spawned a vast literature, it has been argued that the theoretical integration of empirical findings is still limited (e.g., Monroe & Read, 2008). In particular, a realistic formalized theoretical framework is lacking that can be directly related to empirical data through statistical estimation and fitting techniques. In the current chapter, we put forward such a formalized measurement model of attitudes, and argue that this model shows promise in integrating our understanding of the structural and dynamical properties of attitudes.

Any formal measurement model of attitudes needs to fulfill two basic properties. First, it must address how multiple responses on an attitude questionnaire relate to the attitude construct. Second, it must provide an explanation of the correlations among these multiple responses. Historically the most influential model of attitudes that fulfills these two basic properties has been the tripartite model of attitudes. In this model, attitudes are assumed to consist of cognitive, affective, and behavioral components (e.g., Bagozzi, Tybout, Craig, & Sternthal, 1979; Breckler, 1984; Eagly & Chaiken, 1993; Fishbein & Ajzen, 1975; Rosenberg et al., 1960). Formalized accounts of the tripartite model assume that attitudes act as latent variables that cause these three components, which in turn cause specific responses to attitude questions. Due to a number of problems discussed below this view of attitudes has fallen out of vogue (e.g., Fazio & Olson, 2003b; Zanna & Rempel, 1988). However, no dominant alternative formalized measurement model of attitudes has yet replaced it. To fill this gap, the present chapter presents a formalized measurement model of attitudes based around network theoretical principles.

In evaluating the tripartite model's contributions to attitude research, it is important to distinguish between how the model *describes* and *explains* attitude structure. In the descriptive sense, the tripartite model represents the basic correlational structure of responses to attitude questions: Different responses to the same attitude object are substantively interrelated. However, in the explanatory sense, the tripartite model is limited because it explains the correlations among responses to attitude questions in terms of the shared influence of a small number of latent variables. As discussed below, it is our view that such latent variable models do not offer plausible representations of the structure of attitudes in relation to formation and dynamics.

Recently developed connectionist models of attitudes, on the other hand, may better capture the dynamics of attitude formation and change (e.g., Monroe & Read, 2008; van Overwalle & Siebler, 2005). In these connectionist models, human cognition is simulated with a network of several interrelated nodes. These models provide plausible mechanistic explanations of how attitudes form and change as a result of the interplay between evaluative reactions that concern the attitude object. However, a weakness of current connectionist models of attitudes is that they are separated from the dominant statistical approaches used in the analysis of attitude data, and their relevance to empirical phenomena is thus indirect. For instance, while connectionist models are suited to address the inherent complexity of attitudes, currently implemented models rely solely on simulated data (Monroe & Read, 2008). Therefore, their description of attitudinal processes tends to be at the metaphorical level. Thus, while the tripartite model can be fitted to empirical data but relies on a substantively implausible model of attitudes, current connectionist models of attitudes provide a plausible account of attitude formation and change but cannot be fitted to empirical data.

Here we propose a new network model of attitudes that can address both the above concerns: the Causal Attitude Network (CAN) model. This model conceptualizes attitudes as networks that consist of evaluative reactions and interactions between these reactions. Relevant reactions include beliefs, feelings and behaviors toward an attitude object. Interactions between these reactions arise through direct causal influences (e.g., the belief that snakes are dangerous *causes* fear of snakes) and mechanisms that support evaluative consistency between related contents of evaluative reactions (e.g., people tend to align their belief that snakes are useful with their belief that snakes help maintain ecological balance). In the CAN model, the structure of attitude networks is held to conform to a small-world structure (Watts & Strogatz, 1998): evaluative reactions that are similar to each other form tight clusters, which are connected by a sparser set of 'shortcuts' between them. Importantly, the CAN model allows for the application of *empirical* network models (i.e., models in which observed variables are treated as causally related nodes in a network; Schmittmann et al., 2013; van der Maas et al., 2006). Such models can be fitted to actual data, but can also deal with complex systems (Barabási, 2011) such as attitudes, and thus combine the explanatory power of connectionist approaches with empirical analyses of attitude structure. This combination allows us to derive a realistic psychometric conceptualization of attitudes.

The outline of this chapter is as follows. First, we discuss the current lack of formalized measurement models in the attitude literature. Second, we combine the notion of cognitive consistency, recent connectionist modeling of attitudes (Monroe & Read, 2008) and recent advancements in applying network theory in psychology (e.g., Cramer, Waldorp, van der Maas, & Borsboom, 2010; for excellent discussions of the relevance of network analysis to the social sciences in general and psychology in particular see Borgatti, Mehra, Brass, & Labianca, 2009; Westaby, Pfaff, & Redding, 2014) to derive a set of requirements for a realistic formalized measurement model of attitudes. Third, based on these requirements we develop the CAN model and discuss the proposed small-world structure of attitudes that underlies it. Fourth, we discuss the CAN model's perspective on attitude formation and structure, attitude stability and change, and attitude strength. Fifth, we discuss possible extensions of the CAN model, the model's implications for the assessment of attitudes, and some possible avenues for further study of the CAN model.

2.1.1 The Need for a Formalized Measurement Model of Attitudes

Attitude research was one of the first fields in psychology in which researchers developed and tested formalized measurement models (e.g., Bagozzi, 1981; Bagozzi & Burnkrant, 1979; Breckler, 1984). These models rested on the tripartite model of attitudes, which, as we noted, holds that attitudes consist of affective, behavioral and cognitive components (e.g., Eagly & Chaiken, 1993; Fishbein & Ajzen, 1975; Rosenberg et al., 1960) and treated the components as unobservable common causes of observable variables (i.e., reflective latent variables). This conceptualization of attitudes, however, has been criticized for not being able to integrate inconsistencies between attitude and behavior, as the model assumes that behavior is part of the attitude (e.g., Cacioppo, Petty, & Green, 1989; Fazio & Olson, 2003a; Tesser & Shaffer, 1990; Zanna & Rempel, 1988). Based principally on this shortcoming, the formalized account of the tripartite model has largely fallen out of vogue despite the lack of an acceptable formalized measurement model to replace it.

Due to this omission, currently there is no satisfactory explanation for a pervasive finding across the attitude field: attitude items display a positive manifold (i.e., attitude items of the same valence are substantively positively related and attitude items of different valence are substantively negatively related; Bagozzi et al., 1979; Bagozzi & Burnkrant, 1979; Breckler, 1984; Conner, Godin, Sheeran, & Germain, 2013; Haddock, Zanna, & Esses, 1993, 1971; Ostrom, 1969; van den Berg, Manstead, van der Pligt, & Wigboldus, 2005). This positive manifold is reflected in the findings that: (1) items, which assess the same component, are highly interrelated (e.g., judging snakes as dangerous, ugly); and (2) the fitted latent variables representing the components are substantively interrelated (e.g., beliefs and feelings toward snakes).

The formalized account of the tripartite model appeared to provide a parsimonious explanation of the interrelations of items and variables assessing the same component — such items are related because they depend on the same underlying component. However, adding to the former discussed critique of the tripartite model, this explanation is likely to be incorrect because it rests on two assumptions of latent variable modeling that are improbable in the context of attitudes. First, items, or indicators, must be *locally independent* and second, items must be *exchangeable* (e.g., Bollen, 1989; Borsboom, 2008; Borsboom, Mellenbergh, & van Heerden, 2003; Cramer et al., 2010; Jöreskog, 1971).

Local independence refers to the assumption that indicators, which measure the same latent variable, have no direct causal influence on each other (e.g., two thermometers are independent when temperature is held constant; Borsboom, 2008). In the context of attitudes, this assumption is unrealistic, as it would mean that there are no interactions between evaluative reactions. As discussed later, this assumption is at odds with the notion of cognitive consistency and with the basic mechanisms of recent connectionist models of attitudes.

Exchangeability of indicators refers to the assumption that adding indicators to a questionnaire only increases reliability but does not add substantial information (e.g., adding thermometers to a perfect thermometer is superfluous; Bollen & Lennox, 1991). If a value of a person on any one of the observed indicators were measured perfectly without any error (i.e., the true score), this value would exhaustively characterize that person's position on the latent variable; thus, as soon as the true score on any one indicator is known, the other indicators cannot add independent information. In other words, the relations between true indicator scores are deterministic: The reflective latent variable model can in fact be derived from the assumption that all true indicator scores are perfectly correlated (Jöreskog, 1971). In the context of attitudes, this would mean that if a given evaluative reaction (e.g., a given belief toward a presidential candidate) truly changes (i.e., change caused by the underlying latent variable), the other evaluative reactions, which belong to the same attitude component (e.g., all other beliefs toward the presidential candidate), have to change the exact same amount. All other variation is attributed to random error. In our view, such an account of attitudes is much too restrictive and not in accordance with views that emphasize internal inconsistencies within attitudes (e.g., Newby-Clark, McGregor, & Zanna, 2002; Thompson, Zanna, & Griffin, 1995; van Harreveld, van der Pligt, & de Liver, 2009).

If one is to reject the idea that beliefs, feelings and behaviors regarding an attitude object reflect underlying attitude components, the question arises as to how such evaluative reactions relate to the attitude concept. The currently prevalent view holds that beliefs, feelings and behaviors represent different classes of channels through which attitudes are formed (e.g., Zanna & Rempel, 1988).[1] As Fabrigar, MacDonald, and Wegener (2005) state:

> ...the contemporary view holds that an attitude is an entity distinguishable from the classes of affect, behavior, and cognition. An attitude, therefore, does not consist of these elements, but is instead a general evaluative summary of the information derived from these bases (Cacioppo et al., 1989; Crites, Fabrigar, & Petty, 1994; Zanna & Rempel, 1988). (p. 82)

This perspective on attitudes also forms the basis for Fazio's (1995, 2007) influential account of attitudes, in which attitudes are defined as associations between an attitude object and a summary evaluation. According to Fazio (1995, 2007), this summary evaluation is derived from cognitive, affective, and/or behavioral information.

Conversely, the view that attitudes are *causes* of cognition, affect and behavior can also be found in current theorizing. This view is, for example, expressed by Eagly and Chaiken (2007), who state that "Attitude is ...a tendency or latent property of the person that gives rise to judgments as well as to many other types of responses such as emotions and overt behaviors." (p. 586)

To summarize, current theorizing on the relation between evaluative reactions and attitudes holds that attitudes can be formed by cognition, affect, and behavior and that attitudes in turn also influence cognition, affect, and behavior. A formal model of attitudes should thus integrate

[1] This view on attitudes is more related to formative factor models, in which the indicators cause the construct (e.g., Bollen & Lennox, 1991). We refrain from discussing such models in detail, because formative factor models do not make any assumptions on the correlational pattern between the measurements (e.g., Schmittmann et al., 2013) and are therefore of limited use as measurement models.

this bidirectional influence between the components of attitudes and the attitude itself. Such a model should furthermore integrate the pervasive finding that evaluative reactions are generally substantively positively interrelated (e.g., Bagozzi & Burnkrant, 1979; Breckler, 1984; Conner et al., 2013; Haddock et al., 1993, 1971; Ostrom, 1969; van den Berg et al., 2005).

2.1.2 From Cognitive Consistency to Network Models

In this section, we discuss the notion of cognitive consistency and use this notion to derive a formalized measurement model of attitudes. In contrast to the formalized conceptualization of the tripartite model of attitudes, which assumes that correlations between evaluative reactions are spurious, we propose that the correlations between evaluative reactions are meaningful because they stem from direct interactions between the evaluative reactions. Through these interactions the attitude construct is formed.

The notion of cognitive consistency plays a central role in both classic and contemporary theorizing in social psychology (Gawronski, 2012; Gawronski & Strack, 2012). Both Heider's (1946, 1958) balance theory and Festinger's (1957) cognitive dissonance theory make the assumption that humans have a basic need for consistency between their cognitions. Consistency theories also extend to the notion that feelings and beliefs should also be at least somewhat consistent (i.e., affective-cognitive consistency; Rosenberg et al., 1960) and that people are motivated to reduce inconsistencies within their attitudes (e.g. van Harreveld, van der Pligt, & de Liver, 2009). Consistency theories thus assume that evaluative reactions have a tendency to align with each other.

Recently, the mechanism of parallel constraint satisfaction was used to implement a formalized conceptualization of cognitive consistency in a connectionist model of attitudes (Monroe & Read, 2008). Parallel constraint satisfaction in the context of attitudes holds that beliefs impose constraints on other beliefs. For example, if two beliefs are positively connected and one of the beliefs becomes activated, the other belief is likely to become activated as well. On the other hand, if two beliefs are negatively connected and one belief becomes activated, activation of the other belief is subsequently inhibited. Whenever the attitude object is activated in working memory, the associated beliefs self-organize in such a way that the constraints are more and more satisfied (i.e., become more consistent). The system thus strives for an optimized representation of the attitude object.

Using simulation studies, it was shown that the model built on the mechanism of constraint satisfaction can account for several well-known empirical findings on attitudes (e.g., thought-induced attitude polarization, systematic versus heuristic processing, implicit attitude processes; Monroe & Read, 2008). In addition to the insights from cognitive consistency theories, parallel constraint satisfaction further establishes that the need for consistency is a driving factor in attitude formation and change (see also Holyoak & Simon, 1999; Simon & Holyoak, 2002; Simon, Snow, & Read, 2004; Simon, Krawczyk, & Holyoak, 2004; Shultz & Lepper, 1996).

To derive a measurement model of attitudes that integrates the above-discussed notions of consistency and optimization, we use *empirical* network models that have recently been applied to research on clinical disorders and personality (Borsboom & Cramer, 2013; Cramer et al., 2010, 2012). In empirical network models, relations between observed variables are not assumed to reflect an underlying factor as, for example, general intelligence, major depression, or in the current discussion, an attitude (Borsboom, 2008; Cramer et al., 2010; van der Maas et al., 2006). Rather, relations between variables are assumed to stem from a network of causally related variables. The relation between, for example, judging snakes as dangerous and as ugly arises through a direct causal connection between the judgments in order to make the attitude more consistent.

A network is a system of such interrelated variables. Some of these variables are directly connected and others only indirectly connected through other variables. While judging snakes as ugly is probably directly related to judging snakes as dangerous, judging snakes as ugly might cause fear of snakes only through also judging snakes as dangerous. As long as the variables in a network are all connected at least indirectly, information can flow from one variable to all other variables in a network. The variables in the network thus all align to some extent (i.e., they become correlated), so that the variables can be regarded as parts of the same system.[2]

A fundamental difference between latent variable models and empirical network models concerns the causal power of observable variables. In latent variable models, observable variables have virtually no causal power, as all causation flows from the latent variable to the observable variables. Empirical network models allow for more causal power of ob-

[2]One might object that alignment of variables suggests some exchangeability between the variables. To the extent that they share variance, this objection is correct. However, variables seldom perfectly align and are therefore also not completely exchangeable in terms of shared variance. Furthermore, even if two variables perfectly align for some time, this does not exclude the possibility that they are influenced by different factors at other times.

servable variables. From a network perspective, observable variables are not merely indicators of a psychological construct — the construct is isomorphic to the observable variables and their causal connections. From this perspective, the attitude construct is isomorphic to evaluative reactions toward a given attitude object and the interactions among those evaluative reactions (i.e., the whole network of the evaluative reactions and the interactions between these reactions represents the attitude construct). It then follows that observable variables are part of the construct and therefore can influence the construct (i.e., the whole system of the causally connected variables) and the construct can also influence the observable variables. For example, judging a snake as dangerous influences the system of the attitude toward snakes through its causal connections to the other evaluative reactions but is also influenced by the system of the attitude through these causal connections. Judging snakes as dangerous can thus both (to some extent) be a cause of a negative attitude toward snakes and also a result or expression of a negative attitude. Network models are furthermore equally well suited to explain interrelations between observable variables as latent variable models are. Several simulations that aimed to explain the positive manifold of intelligence subtests with a network model showed that networks can shape data that resembles different factor-analytic findings (van der Maas et al., 2006).

By combining insights from cognitive consistency theories, connectionist models of attitudes and empirical network models, we derive two main requirements for a measurement model of attitudes. First, correlations between evaluative reactions stem from pairwise interactions between the reactions and, second, these interactions are aimed at optimization of the consistency of the evaluative reactions. These two requirements are readily fulfilled by the recently proposed application of the Ising model (Ising, 1925) to psychological data (Epskamp, Maris, Waldorp, & Borsboom, 2018; van Borkulo et al., 2014). The Ising model belongs to the class of Markov Random Field models used in network analysis and constitutes the Markov Random Field for binary data (Kindermann & Snell, 1980). Originating in statistical physics, the Ising model has been applied in several research areas. The interested reader is referred to Epskamp, Maris, et al. (2018) for a thorough discussion of the Ising model.

In the Ising model, nodes can be of two states (-1,1) (Kindermann & Snell, 1980). These nodes can represent all kinds of objects that can be in two states (van Borkulo et al., 2014). For example, dipoles of a magnet that are either in "spin up" or "spin down", neurons that either fire or not fire, psychopathological symptoms that are either present or absent, or evaluative reactions that are endorsed or that are not endorsed. Each node in a network is connected to a given number of neighboring nodes and these connections can be either positive or negative. If the connection between two nodes is positive then two neighboring nodes will be pressured to assume the same state and if the connection is negative then the two nodes will be pressured to assume the opposite states. The connections can also vary in strength. Because of the connections between the nodes, the first requirement of a measurement model of attitudes — pairwise interactions between evaluative reactions — is fulfilled.

The second requirement of a measurement model of attitudes — optimization of consistency — is fulfilled by the axiom of the Ising model that the system strives to reduce energy expenditure. In a given Ising model, configurations of the system that cost a lot of energy to maintain are less likely than configurations that cost less energy. Inconsistent attitudes therefore cause much energy expenditure and the system will strive to reduce this inconsistency. Given two positively connected nodes, the configurations in which both nodes are either positive (+1, +1) or negative (-1, -1) are thus more likely than the configurations in which the nodes are not aligned (+1, -1), (-1, +1), as the latter configuration costs more energy to maintain. The extent to which the unaligned configurations cause energy expenditure depends on the strength of the connection between the nodes.

To summarize, the Ising model represents a promising conceptualization of attitudes as this conceptualization readily integrates both interactions between evaluative reactions and the need to maximize cognitive consistency. In the next section, we present the specifics of the Causal Attitude Network (CAN) that implements this conceptualization.

2.2 The Causal Attitude Network (CAN) Model

In the CAN model, attitudes are conceptualized as networks of interacting evaluative reactions (e.g., feelings, beliefs and behaviors toward an attitude object) and the dynamics of the networks conform to the Ising model. Evaluative reactions are represented as *nodes* in the networks and causal influences between these reactions are represented as *edges* (i.e., links between the nodes). Edges can represent either excitatory or inhibitory influence and can have varying weights (i.e., the causal influence between evaluative reactions varies). Attitude networks strive for an optimized consistent representation of the attitude object to reduce energy expenditure. To acquire a consistent state, evaluative reactions of the same valence generally have excitatory influence between them and evaluative reactions of different valence generally have inhibitory influence between them.

However, as individuals are also motivated to hold at least somewhat accurate attitudes (e.g., Petty & Cacioppo, 1986; Chaiken et al., 1989), optimization is bound by the motivation to have an accurate attitude. While striving only for consistency would lead to perfectly aligned evaluative reactions, striving only for accuracy can, in some instances, lead to completely unaligned evaluative reactions. To deal with this trade-off between optimization of consistency and accuracy, attitude networks are proposed to show clustering (i.e., different sets of evaluative reactions are highly interconnected). Clustering allows for energy reduction within clusters (e.g., all evaluative reactions toward a person that pertain to the dimension of warmth are highly aligned) but also allows for accuracy by having unaligned or even misaligned clusters that do not cost much energy (e.g., the evaluative reactions that pertain to the dimension of warmth are not highly aligned to the evaluative reactions that pertain to the dimension of competence).

The trade-off between optimization of consistency and accuracy is somewhat reminiscent of the trade-off between effort and accuracy in decision-making (e.g., J. W. Payne, Bettman, & Johnson, 1988, 1993). Noteworthy from the line of research on the trade-off between effort and accuracy is that in many scenarios decision-strategies that do not require much effort (e.g., heuristics) fare equally well or even better than decision-strategies that require much effort (e.g., J. W. Payne et al., 1988; Gigerenzer, Todd, & the ABC Research Group, 1999). This might indicate that also in the case of attitudes, a representation that optimizes the trade-off between consistency and accuracy might be the most adaptive representation.

2.2.1 Attitude Formation and Structure

From the perspective of the CAN model, attitudes start out with one or just a few specific evaluative reactions. These first reactions then serve as a model for the person to predict which other characteristics the attitude object might have. This is in accordance with the free-energy principle, which holds that humans need to make inferences from learned information to derive predictions (Friston, 2009; Friston, Daunizeau, Kilner, & Kiebel, 2010). Then, also in line with the free-energy principle, information is sought that conforms to this prediction (see also Hart et al., 2009 for a meta-analysis that indicates that individuals generally prefer information that supports their attitudes). Evaluative reactions therefore cause readiness of other evaluative reactions. However, this readiness is not deterministic as it is also possible that no confirming information is found. Such a situation is likely to arise when individuals are highly motivated to be accurate, which lowers their preference for information that supports their attitudes (Hart et al., 2009).

As individuals are motivated to make correct inferences, it is likely that the strength of the readiness depends on the similarity of the reactions because correct inferences are more likely to be made for similar evaluative reactions. For example, two judgments that pertain to the dimension of warmth (e.g., friendly and sincere) would cause more readiness of each other than two judgments that pertain to different dimensions (e.g., friendly and competent; cf. Fiske, Cuddy, Glick, & Xu, 2002). Reactions from different attitudinal components are also generally less similar than reactions from the same component and are therefore also generally less closely connected. Another factor that makes it more likely that an evaluative reaction causes readiness for another evaluative reaction is that the evaluations share the same valence because evaluations of differing valence can behave relatively independent of each other (Cacioppo & Berntson, 1994). Connecting this line of reasoning to research on the development of networks, a recent study showed that nodes that make new connections are more likely to connect to similar nodes than to less similar nodes (e.g., a US webpage is more likely to connect to another US webpage than to a Russian webpage; Papadopoulos, Kitsak, Serrano, Boguña, & Krioukov, 2012).

Another factor of importance in the attachment of new nodes to the network is the popularity of the nodes (i.e., how many connections a node already has; Barabási & Albert, 1999). This means that nodes are more likely to connect to nodes that already have many connections (a phenomenon known as preferential attachment). In the case of attitudes, this would mean that evaluative reactions that already have many connections are more likely to lead to the activation of additional evaluative reactions. The proposed mechanism behind this effect is that evaluative reactions that are strongly connected already have proven to be predictive in the *past*, which makes such evaluative reactions more likely to cause readiness of other evaluative reactions in the *present*.

Let us illustrate how attitude networks might take shape using the example of someone, say Bob, developing a positive attitude toward Barack Obama during the American presidential election campaign of 2008. First, Bob formed the impression of Obama being an honest per-

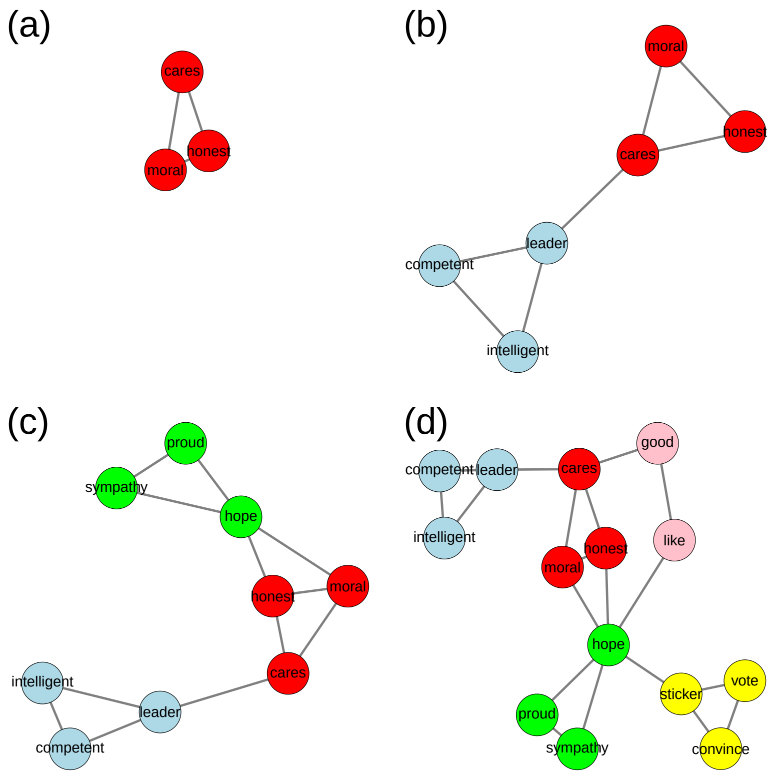


Figure 2.1: Hypothetical attitude network at four points in time. Nodes represent evaluative reactions and edges represent causal influence between the evaluative reactions. Note that for reasons of simplification, all edges represent excitatory influence and all edges have the same strength. The layouts of these graphs are based on the Fruchterman-Rheingold algorithm (Fruchterman & Reingold, 1991), which places closely connected nodes near each other. As all networks shown in this chapter, this network was created using the R-package qgraph (Epskamp et al., 2012).

son. Quickly after forming this initial simple impression, Bob also formed similar judgments about Obama, like judging him as a moral person who cares about people like Bob. The formed judgments hold each other in check, so that none of these judgments can readily change without inflicting some change on the other judgments. This already constitutes a small network, see Figure 2.1a.

After Bob has thought a bit more about Obama, he also extended his judgment to other dimensions and came to the conclusion that Obama is also intelligent and competent and that he will probably be a good leader. These judgments form a new cluster but are also to some extent linked to the other judgments through the connection between judging Obama as a good leader and judging him to care about people like Bob, as Bob thinks that this is a crucial aspect of being a good leader. The attitude network has thus grown and now consists of two *clusters* (i.e., set of nodes that are highly interconnected), which are connected by a *shortcut*, see Figure 2.1b. The reason that this edge represents a shortcut is that removing this edge would substantially decrease the global connectivity (i.e., average connectivity of each node with all other nodes; West, 1996) of Bob's attitude network (Watts, 1999). In the current example, removing the edge between judging Obama as caring and judging Obama to be a good leader would result in Bob's attitude network no longer being fully connected.

At some point, Bob also had evaluative reactions of a more affective nature toward Obama. Because he judged Obama as honest and moral, he also started to feel hopeful toward Obama and this in turn caused him to feel proudness and sympathy toward him, see Figure 2.1c. Again, these different affective reactions cannot change without exerting some pressure to change on the other affective reactions. Furthermore, Bob's feeling of hope toward Obama and his judgments that Obama is honest and moral are closely connected, so that when one of these evaluative reactions increases or decreases, the other reactions will also more readily increase or decrease.

Due to his already very positive attitude, Bob also started to convince other people to vote for Obama, he placed a sticker on his car saying 'Vote Obama' and, of course, at the election he voted for Obama. From the more specific evaluative reactions toward Obama, more general evaluations subsequently emerged. For example, Bob would state that he likes Obama and that he generally sees him as a good person. These new clusters attached to the evaluative reactions of judging Obama as caring and feeling hopeful toward Obama due to the popularity (i.e., number of connections) of these evaluative reactions, see Figure 2.1d.

As the above example illustrates, attitude networks are expected to show a structure with high clustering, in which these clusters are connected through shortcuts. Through these shortcuts, attitude networks have high global *connectivity* (i.e., all nodes on average are closely connected to each other). The combination of high clustering and high connectivity is known as a small-world structure (Albert & Barabási, 2002; Watts & Strogatz, 1998). The formalized definition of small-world networks holds that such networks have higher clustering than a random graph (i.e., network in which the nodes are randomly connected) but about the same connectivity as a random graph. For example, the network shown in Figure 2.1d has a clustering index almost three times as high (C = .59) as the clustering index of a corresponding random graph (C = .21) but only a slightly higher average shortest path length (L = 2.74) than the corresponding random graph (L = 2.33), implying slightly lower connectivity. For all network-based calculations in this chapter, we used either the R-package qgraph (Epskamp et al., 2012) or the R-package igraph (Csárdi & Nepusz, 2006).

The small-world structure has been observed for several networks from distinct areas of research. For example, small-world structures have been found for the power grid of the western United States; collaboration between film actors; neural networks of worms (Watts & Strogatz, 1998), monkeys (e.g., Stephan et al., 2000), and humans (e.g., Achard, Salvador, Whitcher, Suckling, & Bullmore, 2006); human language (Ferrer i Cancho & Solé, 2001); and symptoms of psychological disorders (Borsboom, Cramer, Schmittmann, Epskamp, & Waldorp, 2011).

To provide a first test of the hypothesis that attitude networks have a small-world structure, we used the American National Election Data (ANES) of 1984 (see Appendix A for a detailed description of the data set). In the ANES of 1984, evaluative reactions toward the presidential candidates were assessed in a nationwide random sample of 2,257 participants. Participants were asked whether or not they attributed several positive characteristics to each candidate (e.g., whether the candidate is a decent, intelligent, or moral person) and whether they had ever had positive or negative feelings toward each candidate (e.g., feelings of hope or anger). We used the participants' responses toward these evaluative reactions to estimate attitude networks for the attitudes toward each presidential candidate.[3]

To estimate attitude networks, we used the eLasso-procedure, which is designed to find the optimal Ising model for a set of data by regressing each variable on all other variables (van Borkulo et al., 2014). The regression function is subjected to regularization to control the size of the statistical problem (see Friedman, Hastie, & Tibshirani, 2008; Tibshirani, 1996). For each node, the set of edges that displays the best fit to the data is selected based on the fit of the regression functions according to the

[3]The reason that we analyzed the ANES of 1984 instead of any other ANES is that the number of assessed evaluative reactions was the largest in the ANES of 1984. As the evaluative reactions are represented as nodes, the size of the network (i.e., number of nodes) depends on the number of assessed evaluative reactions. As the small-world index linearly increases with the size of the network in small-world networks (Dunne, Williams, & Martinez, 2002; Humphries & Gurney, 2008), the power to detect a small-world structure is higher for the ANES of 1984 than for any other ANES. Note, however, that the size of the estimated networks based on the ANES of 1984 is still relatively small.

Extended Bayesian Information Criterion (Chen & Chen, 2008). Weights of the edges are then based on the regression parameters in the selected neighborhood functions (van Borkulo et al., 2014).[4]

We then calculated the small-world index for the unweighted networks (Humphries & Gurney, 2008). A small-world index higher than one indicates that the network has a small-world structure. To test whether the small-world index was significantly higher than one, we calculated confidence intervals using 1000 Monte-Carlo simulations of random graphs (Humphries & Gurney, 2008).

The estimated networks are shown in Figure 2.2. The network of the attitude toward Ronald Reagan had a small-world index of 1.16 and the upper limit of the 99.9% confidence interval for the corresponding random graphs was 1.13 ($C_{attitude\ network}$ = .62, $L_{attitude\ network}$ = 1.47, $C_{random\ network}$ = .53, $L_{random\ network}$ = 1.47). The network of the attitude toward Walter Mondale had a small-world index of 1.25 and the upper limit of the 99.9% confidence interval for the corresponding random graphs was 1.19 ($C_{attitude\ network}$ = .57, $L_{attitude\ network}$ = 1.55, $C_{random\ network}$ = .46, $L_{random\ network}$ = 1.54). Both networks thus had a small-world structure. Apart from the global structure, it can also be seen that similar evaluative reactions are more closely connected than dissimilar evaluative reactions. For example, both positive and negative feelings form distinct clusters (see Figure 2.2). Also, judgments that pertain more to the warmth-dimension (e.g., the candidate is fair, cares about people and is compassionate) and judgments that pertain to the competence-dimension (e.g., the candidate is intelligent, knowledgeable and hard-working) are each closely connected to each other.

To summarize, the CAN model holds that evaluative reactions cause readiness of related evaluative reactions to the same attitude object and through this process attitude networks take shape. Similar evaluative reactions tend to cluster and these clusters are connected by shortcuts, which give rise to the small-world network structure of attitudes. Having

[4]Note that also other methods than the eLasso-procedure can be used to estimate networks from psychometric data (e.g., Costantini et al., 2015). Two common practices involve estimating edges by using zero-order correlations or using partial correlation between any given pair of the nodes in the network while controlling for all other nodes (e.g., Cramer et al., 2010, 2012). The eLasso-procedure, however, has clear advantages compared to these procedures. First, the eLasso-procedure provides an estimation of the causal structure underlying the data, which zero-order correlations do not. Second, the eLasso-procedure produces less false positives than does using partial correlations (Costantini et al., 2015). The reason for that is that the eLasso-procedure uses regularization, which reduces the number of false positives. In fact, simulation studies have shown that the eLasso-procedure has very high specificity (van Borkulo et al., 2014).

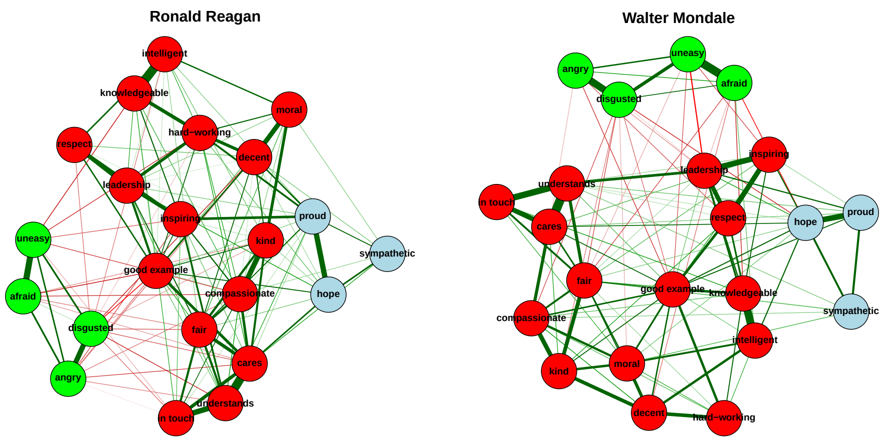


Figure 2.2: Estimated attitude networks toward the two main candidates in the US presidential election in 1984. Red nodes represent positive judgments, blue nodes represent positive feelings and green nodes represent negative feelings. Green edges indicate excitatory influence between the nodes and red edges indicate inhibitory influence between the nodes. Thicker edges represent higher weights of the edges. The same algorithm as for Figure 2.1 was used for the layout of these graphs.

discussed the formation and structure of attitude networks we now turn to the implications of the CAN model for change and stability in formed attitudes.

2.2.2 Attitude Stability and Change

After an attitude network has formed, the CAN model suggests that different evaluative reactions tend to hold each other in check. While in the formation phase of an attitude, the influence of evaluative reactions on other evaluative reactions is unidirectional, evaluative reactions in formed attitude networks probably have bidirectional influence on each other. The reason for this is that an evaluative reaction B that was activated by another evaluative reaction A will also cause further readiness in evaluative reaction A. As an example let us return to Bob's attitude network. After having formed the judgment that Obama cares for people like Bob *because* he earlier judged Obama to be honest, judging Obama as caring also serves as a model of how honest Obama is. So, if Bob observes an incident that strengthens his judgment of Obama as caring, his judgment of Obama being honest will also strengthen to some extent.

While we have discussed attitude formation and the behavior of formed attitudes in isolation, the distinction between these two phases of attitudes will not be that clear-cut in reality. New nodes and new connections will probably be formed continuously and the extent of this likely depends on how often a person interacts in some way with the attitude object. We discuss this point further in the section on the relation between attitude strength and network connectivity.

Apart from implications for attitude stability, conceptualizing attitudes as networks also has implications for attitude change. The most straightforward implication of conceptualizing attitudes as networks for attitude change is that attitudes can be changed via a plethora of different processes as each node in the attitude network can serve as a gateway to instigate change in the network. Looking back at Figure 2.1d, change in the network could, for example, be instigated by cognitive dissonance (e.g., Bob did not vote for Obama because of minor situational constraints and because of this a more negative evaluation spreads through the network; cf., Festinger, 1957), evaluative conditioning (e.g., pairing Obama with images related to hope through which an even more positive evaluation spreads through the network; cf., De Houwer, Thomas, & Baeyens, 2001; Jones, Olson, & Fazio, 2010) or by presenting arguments (e.g., a friend convinces Bob that Obama is not so competent after all from which again a more negative evaluation spreads through the network).

Whether the state of the targeted evaluative reaction changes (e.g., from +1 to -1), however, is a function of not only the strength of external pressure but also of the state of the neighboring nodes and the strength of the links between the targeted node and the neighboring nodes. If a single node changes to a state that is incongruent to its links with neighboring nodes, the energy expenditure of the system rises. The amount of energy expenditure of a given configuration is calculated using the Hamiltonian function :

$$H(x) = -\sum_{i} \tau_i x_i - \sum_{i,j} \omega_{ij} x_i x_j, \quad (2.1)$$

where τ is the threshold of any given evaluative reaction x_i and represents the disposition of the given evaluative reaction to be endorsed (1) or not endorsed (-1). A threshold higher than 0 indicates that the probability of the given evaluative reaction to be endorsed is higher (given the absence of any influence of neighboring evaluative reactions) than the probability that the evaluative reaction is not endorsed. Conversely, a threshold lower than 0 indicates that the probability of the given evaluative reaction to be endorsed is lower than the probability that the evaluative reaction is not endorsed.

ω_{ij} is the weight of the interaction between x_i and its neighboring evaluative reaction x_j. A weight higher than 0 indicates that that the interaction between two evaluative reactions is positive (e.g., if one evaluative reaction is endorsed, the probability that the other evaluative reaction is endorsed heightens). Conversely, a weight lower than 0 indicates that that the interaction between two evaluative reactions is negative (e.g., if one evaluative reaction is endorsed, the probability that the other evaluative reaction is endorsed lowers).

The probability of a given configuration can be calculated with the following equation:

$$Pr(X = x) = \frac{1}{Z} exp(-H(x)), \tag{2.2}$$

where Z is the normalizing constant that guarantees that the probabilities sum to 1 and is given by:

$$Z = \sum_{x} exp(-H(x)) \tag{2.3}$$

To acquire a consistent state, weights between the evaluative reactions of the same valence (e.g., honest and fair) have to be positive and weights between evaluative reactions of different valence have to be negative (e.g., honest and mean). The strength of the weights is a function of the amount of interaction with the attitude object (Monroe & Read, 2008) and the similarity of the evaluative reactions. The CAN model holds that the threshold depends on several factors, implying that evaluative reactions differ in their disposition to be endorsed. While a high threshold indicates that a given evaluative reaction has a strong disposition to be endorsed, a low threshold indicates that a given evaluative reaction has only a weak disposition to be endorsed (see Equation 2.1). First, some evaluative reactions are likely to have inherently higher thresholds than other evaluative reactions. For example, some emotions are experienced more frequently (e.g., joy, anger) than others (e.g., euphoria, contempt; Schimmack & Diener, 1997) and are therefore likely to have higher thresholds. Second, thresholds of different persons probably also vary. For example, persons differ in their disposition to adopt negative or positive attitudes (Eschleman, Bowling, & Judge, 2015; Hepler & Albarracín, 2013). Third, whether one has endorsed (not endorsed) an evaluative reaction in the past will probably also heighten (lower) the threshold of this evaluative reaction in the future, implying that the evaluative reaction is more likely to be endorsed (not endorsed) in the future. This postulate is based on the finding that rehearsal of attitudes results in the strengthening of attitudes (e.g., Fazio, 1995). Fourth, external persuasion attempts might also affect the threshold of a given evaluative reaction. One route to change an attitude would thus be to change the threshold of evaluative reactions.

Let us return to Bob to illustrate how an attitude network would be affected if the threshold of an evaluative reaction were to change as a result of a persuasion attempt. For simplification, we will only look at the cluster of the evaluative reactions that Obama is competent, intelligent and a good leader. Let us assume that the thresholds before the persuasion attempt were all equal to .3 and that the weights between all evaluative reactions were all equal to .5. These represent moderately high thresholds and weights (using uniform weights and thresholds is a simplification, as these parameters are likely to vary in real attitude networks, a point we address later).

The persuasion attempt on Bob's attitude was directed at the evaluative reaction that Obama is competent and the attempt was quite strong. Because of this, Bob's threshold of judging Obama as competent dropped to -.3. We can now calculate the energy that each of the configurations will cost and how probable each configuration is. As can be seen in Table 2.1, the configuration in which all three evaluative reactions are endorsed, is the most likely configuration. The most likely scenario for this illustration is thus that, while the persuasion attempt was quite strong, Bob will still think of Obama as competent. This illustrates our point that the success of a persuasion attempt not only depends on the strength of the persuasion attempt but also on the parameters of the attitude network. It is also important to note that it is more probable that evaluative reactions will become uniformly negative after the persuasion attempt, than that the judgment of Obama as competent will change in isolation. This leads to a testable prediction of the CAN model: if one evaluative reaction changes and this change persists, other evaluative reactions are also likely to change.

Changes in the threshold of a given node can be interpreted as rather subtle attitude change. Such changes can, for example, arise if arguments are presented that make one doubt one's current evaluation. However, there might also be situations, in which one not only doubts one's current evaluation but discards it completely. Such attitude change could be modeled by fixing one of the nodes to a given state. This would be the case if, for example, Bob were to receive information which would make it impossible to remain viewing Obama as competent. In this event, only the configurations in which judging Obama as competent is not endorsed feed into Equation 2.2. As a result, the configuration in which none of the pos-

Table 2.1: All Possible States of the Evaluative Reactions Competent, Good Leadership and Intelligent and Their Associated Energies and Probabilities

Competent	Good Leader	Intelligent	$H(x)$	$Pr(X = x)$
-1	-1	-1	-1.2	.24
1	-1	-1	1.4	.02
-1	1	-1	0.2	.06
1	1	-1	0.8	.03
-1	-1	1	0.2	.06
-1	-1	1	0.8	.03
-1	1	1	-0.4	.11
1	1	1	-1.8	.44

Note. Competent has a threshold of -.3 and good leader and intelligent have thresholds of .3, respectively. All weights between the evaluative reactions are equal to .5.

itive evaluative reactions are endorsed becomes the most likely scenario.

While up to this point we have discussed how attitude change takes place in a network, in which all nodes are uniformly connected to each other, nodes in real attitude networks will differ in their connectivity. This can be seen in both the hypothetical attitude network of Bob and the estimated networks of the ANES data. How a given node is connected in the network will influence whether and how change in this node will spread to other nodes. Connecting attitude change to the small-world structure of attitudes, the first important implication of network structure for attitude change is the existence of clusters within the network. As can be seen in Figure 2.2, negative feelings toward the presidential candidates form rather distinct clusters in both attitude networks. If a node in this cluster were to be changed, this change would mostly spread to other nodes in this cluster.

If the different clusters in an attitude network were not strongly connected, configurations of the network, in which clusters do not align to the same evaluation, would cost rather low energy. Because of this, even an attitude network consisting of clusters that are not aligned to each other can be relatively stable. Even misaligned clusters might cause low energy when the weights between the clusters are lower than the inconsistent thresholds of the different clusters. For example, if both negative feelings and positive beliefs in the attitude network toward Ronald Reagan had highly positive thresholds and the (negative) weights between the clusters were low, having both negative feelings and positive beliefs toward Ronald Reagan would not cause much energy expenditure.

Apart from belonging to different clusters, nodes can also differ in their *centrality*. Centrality refers to the structural importance of a given node in the network and three of the most popular centrality measures are *betweenness*, *degree* and *closeness* (Freeman, 1978; Opsahl, Agneessens, & Skvoretz, 2010). Betweenness refers to how often a node lies in the shortest path between two other nodes. The shortest path between two nodes is defined as the shortest distance it takes to 'travel' between two nodes over the edges in the network. The distance is a function of both the number and strength of the edges that lie between two nodes.

Betweenness can be linked to the notion of clustering, as nodes that connect different clusters will generally have high betweenness. If change affects one cluster of the attitude network, whether the change will spread through the whole network depends on the behavior of the nodes that connect this cluster to other parts of the network. In the network of the attitude toward Ronald Reagan, the node with the highest betweenness is the evaluative reaction of whether he sets a good example. As can be seen in Figure 2.2, this node is rather closely connected to the negative affect cluster. Thus, whether change in the negative affect cluster would spread through the network would depend on whether you change your mind that Ronald Reagan sets a good example.

Degree is a function of the number of neighbors a given node has and of the weights of the edges between the given node and its neighbors (Freeman, 1978). Degree thus refers to how strongly a given node is *directly* connected to all other nodes in the network. Closeness is the inverse of the shortest path length between a given node and all other nodes in the network and thus refers to how strongly a given node is both *directly and indirectly* connected to all other nodes in the network. While a node with high degree and/or high closeness is very unlikely to change independent of change in the whole network and vice versa, a node with low degree and low closeness can change rather independently of the other nodes in the network. Therefore, it will be more difficult to change nodes with high degree and/or high closeness than nodes with low degree and low closeness. If change, however, takes place, it will be more consequential if it takes place in a node with high degree and/or closeness than when it takes place in a node with low degree and low closeness. For example, the evaluative reaction with both the highest degree and highest closeness in the network of the attitude toward Ronald Reagan is the judgment of whether he cares about people like oneself. It is thus likely that it would be very difficult to change this judgment but that change in this judgment would affect the attitude network to a large extent.

To summarize, different routes of attitude change can be integrated in the CAN model. First, subtle attitude change can be modeled by decreasing or increasing the threshold of a given node and drastic attitude change can be modeled by fixing a given node to a given state. Attitude change will also depend on the position of the targeted evaluative reaction in the attitude network. Change of an evaluative reaction will first spread to evaluative reactions that belong to the same cluster and the extent to which an attitude network will be affected by change in a single evaluative reaction will depend on the centrality of the evaluative reaction. While highly central evaluative reactions will be likely to resist change, their change will also be more consequential than change in an evaluative reaction that is not central.

2.2.3 Attitude Strength

Strong attitudes are defined by their stability, resistance to change and impact on behavior and information-processing (Krosnick & Petty, 1995; Visser, Bizer, & Krosnick, 2006). Apart from these *key features* of attitude strength, several other *attributes* have been identified that are related to attitude strength, such as extremity (Abelson, 1995), elaboration (Petty, Haugtvedt, & Smith, 1995) and importance (Boninger, Krosnick, Berent, & Fabrigar, 1995). In this section, we show that the CAN model integrates the key features of attitude strength into a single framework. Specifically, in the CAN model, the *global connectivity* (i.e., average shortest path length; West, 1996) of an attitude network can be regarded as a mathematically formalized conceptualization of attitude strength. We first discuss how the global connectivity of a network affects the dynamics of a network and link these differing dynamics to the key features of attitude strength. Then we discuss some empirical support for the proposition that attitude strength and connectivity of attitude networks are related, after which we discuss how connectivity of attitude networks relates to the other attributes related to attitude strength.

As we already hinted at in our discussion of attitude change, stability and resistance to change of a given node in a network depends on how strongly this node is connected to other nodes in the network. By extending this argumentation to the whole network, it follows that nodes in a highly connected network are more stable and resistant than nodes in a weakly connected network and that because of this the whole network can be regarded as more stable and resistant (Kindermann & Snell, 1980; van Borkulo et al., 2014). Note that for these dynamics to occur, the assumption is made that the weights are organized in a way that not much conflicting influence is present (e.g., connections between evaluative reac-

Table 2.2: Configuration of Three Evaluative Reactions (Competent, Good Leader, Intelligent) and Associated Probabilities with Either Congruent or Incongruent Thresholds and Either Low or High Weights

	$Pr(X = x)$			
	$\tau_i\{.3,.3,.3\}$		$\tau_i\{-.3,.3,.3\}$	
Configuration	$\omega_{ij}\{.1,.1,.1\}$	$\omega_{ij}\{1,1,1\}$	$\omega_{ij}\{.1,.1,.1\}$	$\omega_{ij}\{1,1,1\}$
(-1, -1, -1)	.06	.14	.11	.33
(1, -1, -1)	.07	.00	.04	.00
(-1, 1, -1)	.07	.00	.13	.01
(1, 1, -1)	.13	.01	.07	.01
(-1, -1, 1)	.07	.00	.13	.01
(1, -1, 1)	.13	.01	.07	.01
(-1, 1, 1)	.13	.01	.24	.02
(1, 1, 1)	.35	.83	.20	.61

tions of the same valence are mostly excitatory).

To illustrate that highly connected attitude networks are more stable than weakly connected attitude networks, let us return to the example of Bob's judgments that Obama is competent, intelligent and a good leader. To compare the dynamics of these judgments in a highly connected network to the dynamics in a weakly connected network, we can set the weights between the nodes to 1 (representing a highly connected network) or .1 (representing a weakly connected network). To first focus on stability without external pressure, the thresholds were all set to .3. As can be seen in Table 2.2, the probability that all judgments are positive is much higher in the highly connected network than in the weakly connected network. These probabilities can be interpreted as the likelihood that we would observe a given configuration if we measure a person's attitude at a given point in time. So, if we were to measure attitudes of a group of individuals who have a highly connected attitude network at two points in time, we would find a higher correlation between their attitude scores than were we to measure a group of individuals who have a weakly connected network, which mirrors how attitude stability is assessed in the attitude strength literature (e.g., Bassili, 1996; Prislin, 1996).

To illustrate that highly connected attitude networks are also more resistant than weakly connected attitude networks, we return to our example, in which persuasion was targeted at Bob's evaluative reaction that Obama is competent. Again, we set the threshold of this evaluative reaction to -.3 and calculate the probability of what would happen in a highly connected network (all weights equal to 1) and in a weakly connected network (all weights equal to .1). Based on the probabilities shown in Table 2.2, the chance that the targeted evaluative reaction will change is higher for the weakly connected network than for the densely connected network. The chance that the sum score of the three evaluative reactions will reflect a less positive evaluation is also higher for the weakly connected network than for the highly connected network.[5] The probability that the targeted evaluative reaction will become negative is .62 and the expected mean evaluation of the three evaluative reactions is 0.12 for the weakly connected network compared to .38 and 0.28 for the highly connected network. Highly connected attitude networks are thus more likely to resist persuasion attempts than weakly connected networks, which fits what is known about the resistance of strong versus weak attitudes to persuasion attempts (e.g., Bassili, 1996; Visser & Krosnick, 1998).

Impact of attitude on behavior in the attitude strength literature is generally assessed by measuring individuals' attitude at a given point in time and then predicting a related behavior at a later point in time (e.g., Fazio & Williams, 1986; Holland, Verplanken, & van Knippenberg, 2002). There are two reasons to expect attitudes with highly connected networks to be more predictive of behavior than weakly connected networks. First, due to the stability and resistance of highly connected networks, it is more likely that the attitude will be the same at the time it is measured and at the time the behavior is executed.

Second, as we can see in Table 2.2, evaluative reactions in highly connected attitude networks are more likely than evaluative reactions in weakly connected attitude networks to align to each other. An aligned attitude network is likely to be more informative for a decision on whether a related behavior should be executed or not. For example, it would be easier for Bob to decide to vote for Obama if he thinks that Obama is competent, a good leader and intelligent than when he only thinks that Obama is competent and intelligent but not really a good leader. In the latter case, Bob would probably base his decision on other factors (e.g., which of the candidates his friends prefer), which would make his attitude less predictive of his voting behavior. Furthermore, it is possible that the salience of evaluative reactions differs between situations (cf., Sparks, Conner, James, Shepherd, & Povey, 2001). If Bob's evaluative reactions differed in their endorsement, it might be that at the time his attitude is assessed, mostly negative evaluative reactions are salient. By the time he executes the behavior, other evaluative reactions that are mostly positive might be salient.

[5] In the CAN model, the unweighted sum score of the evaluative reactions can be used as a measure of the overall state of the network. A possibility for a more sophisticated calculation for the overall state of the network might be to weigh the evaluative reactions by their closeness centrality, as evaluative reactions with high closeness hold more information about the other evaluative reactions than evaluative reactions with low closeness.

In such a case, his measured attitude would have low predictive value.

The impact of strong attitudes on information processing refers to the power of strong attitudes in directing attention and influencing the way in which incoming information is integrated (e.g., Fazio & Williams, 1986; Houston & Fazio, 1989; Roskos-Ewoldsen & Fazio, 1992). The connectivity of an attitude network influences the way in which incoming information is integrated, because evaluative reactions that are not aligned to each other cost more energy in a highly connected attitude network than in a weakly connected attitude network. There is thus more pressure to fit incoming information to one's evaluative reactions in a highly connected network than in a weakly connected network. This pressure might also lead to heightened attention to attitude objects in order to detect 'attacks' on the attitude network at an early stage.

As we have shown, the key features of attitude strength follow from conceptualizing strong attitudes as highly connected networks. From this reasoning, the hypothesis follows that strong attitudes correspond to highly connected networks. Indeed, this is exactly what we found in a study focusing on attitudes toward presidential candidates in the American presidential elections from 1980-2012 (Dalege, Borsboom, van Harreveld, & van der Maas, 2018b). In this study, we first assigned participants into three strength groups based on whether the participants were interested in presidential campaigns (interest in an attitude object is known to be a reliable indicator of attitude strength; Krosnick, Boninger, Chuang, Berent, & Carnot, 1993). We then checked whether these groups differed in their attitude's stability, extremity and impact on behavior. All of the attributes differed between the groups in the expected direction, so that the high attitude strength group had attitudes with the highest stability, extremity and impact on behavior.

We then estimated networks for each attitude strength group for the attitudes toward the Democratic and Republican candidate at each election and compared the global connectivity of the attitude networks, giving us a total of 18 sets of attitude networks of groups with either low, moderate or high strength. Confirming the hypothesis that strong attitudes correspond to highly connected networks, we found that the groups' networks differed robustly and strongly in their global connectivity. In every set of attitude networks, the network of the strong attitude group had a higher connectivity than average and in all but one set, the network of the weak attitude group had a lower connectivity than average.

Thus, the conceptualization of strong attitudes as highly connected networks directly implies that strong attitudes are stable and resistant. By adding a few simple assumptions, the other two key features of attitude strength — impact on behavior and information processing — can also be integrated in the current framework. Furthermore, initial empirical results show that attitude strength and connectivity of attitude networks are indeed strongly and robustly related. The question then arises of how attributes that are related to attitude strength relate to the connectivity of attitude networks.

In their review on attitude strength-related attributes, Visser et al. (2006) identified elaboration (Petty et al., 1995), importance (Boninger et al., 1995), knowledge (W. Wood, Rhodes, & Biek, 1995), accessibility (Fazio, 1995), certainty (Gross, Holtz, & Miller, 1995), ambivalence (Thompson et al., 1995), structural consistency (Chaiken, Pomerantz, & Giner-Sorolla, 1995), extremity (Abelson, 1995), and intensity (Cantril, 1946) as the most prominent attributes related to attitude strength. It is our view that these attributes can be roughly grouped into three clusters in relation to connectivity of attitude networks: attributes that are determinants of heightened connectivity, attributes that are consequences of connectivity, and attributes that moderate the consequences of connectivity.

Elaboration is probably a direct cause of connectivity. Building on the proposition of connectionist models of attitudes that evaluative reactions self-organize when the attitude object is activated in working memory (Monroe & Read, 2008) and the principle that humans are motivated to decrease free-energy (e.g., Friston, 2009), it follows that connections between evaluative reactions will become stronger when a person elaborates on his or her attitude. The same holds for determinants of attitude importance. Three key determinants of attitude importance are self-interest, social identification and value relevance (Boninger et al., 1995). Self-interest refers to the extent that one perceives that a given attitude has influence on one's life, social identification refers to whether an attitude is relevant to groups a person identifies with, and value relevance refers to how strongly an attitude is linked to a person's value. What these three determinants of attitude importance have in common is that they all make it likely that an attitude will play a large role in a person's life. The attitude object therefore is frequently activated in working memory. This is, for example, illustrated by the finding that individuals often think about attitudes that are important to them (Herzog, 1993; Krosnick et al., 1993). It thus follows that determinants of attitude importance are also likely to determine connectivity of attitude networks.

Knowledge on the attitude object can rather be seen as an amplifier of attitude strength than as being directly related to attitude strength — the effects of attitude strength are more pronounced when there is a large amount of knowledge (e.g., a strong attitude in combination with much knowledge is very resistant to change; W. Wood et al., 1995). In the CAN model, large amounts of knowledge about an attitude object could best be modeled by having a network with many nodes, as it is likely that an individual with much knowledge about an attitude object will also have many different evaluative reactions. Knowledge about an attitude object would thus affect the size of the attitude network but not necessarily the connectivity of the attitude network. The size of a network was shown to amplify the effect of connectivity on network dynamics (Cramer et al., 2016) leading to the hypothesis that attitude networks that are both highly connected and consist of many different evaluative reactions will correspond to stronger attitudes. This is in line with the idea that knowledge amplifies the effects of attitude strength (W. Wood et al., 1995).

In our view, accessibility, certainty, structural consistency, intensity, and extremity are all likely to be caused by connectivity. Accessibility refers to how fast a person can judge whether a given attitude object is positive or negative (e.g., Fazio & Williams, 1986). We would argue that it is easier to judge an attitude object as either positive or negative if one has an attitude that consists of aligned evaluative reactions (which is more likely in a highly connected attitude network) than when the attitude consists of unaligned evaluative reactions (which is more likely in a weakly connected attitude network). This reasoning is supported by the finding that judging an attitude object, to which an individual holds an ambivalent attitude, takes longer than judging an attitude object, to which an individual holds an univalent attitude (Bargh, Chaiken, Govender, & Pratto, 1992; van Harreveld, van der Pligt, de Vries, Wenneker, & Verhue, 2004).

Attitude certainty can be divided into the constructs attitude clarity and attitude correctness (Petrocelli, Tormala, & Rucker, 2007). Attitude clarity refers to how certain a person is what his or her attitude actually is and attitude correctness refers to how convinced a person is that his or her attitude is valid. It is our view that attitude clarity rather than attitude correctness can be directly linked to the connectivity of attitude networks. First, having an aligned attitude network will probably make it more likely that a person is certain about his or her attitude. Second, both attitude clarity and connectivity of attitude networks are likely to be higher when one has frequently interacted with the attitude object. This could also explain why some ambivalent attitudes are held with certainty. This would be the case if, for example, one has received much conflicting information regarding an attitude object and the strongly connected attitude network was not able to settle in a state, in which the evaluative reactions are aligned. We will return to this point in more detail when we discuss the relation between ambivalence and connectivity of attitude networks.

Structural consistency refers to attitudes being evaluative-affective, evaluative-cognitive, and affective-cognitive consistent (Chaiken et al., 1995). Affective-cognitive consistency directly follows from an attitude network being highly connected because the affective and cognitive nodes are likely to align. Also, having an attitude network, in which all (or most) of the evaluative reactions are aligned, makes it likely that both feelings and beliefs will be consistent with one's general evaluation (i.e., evaluating an attitude object as positive or negative overall).

Intensity refers to how strongly an attitude object elicits emotional reactions (Visser et al., 2006). As emotional reactions represent nodes in an attitude network, strong emotional reactions are expected in attitudes with highly connected networks. Returning to Bob's attitude toward Obama, Bob will have strong positive emotional reactions toward Obama when his attitude network is highly connected as these reactions are pressured to align to the overall positivity of the attitude network. In a weakly connected network, his emotional reactions can vary more freely, so that weaker emotional reactions are expected.

Finally, extremity follows directly from high connectivity. As can be seen in Table 2.2, in highly connected attitude networks it is virtually impossible that the sum score of the evaluative reactions takes a moderate value. This finding also links the CAN model to the catastrophe model of attitudes (Flay, 1978; Latané & Nowak, 1994; Zeeman, 1976), which holds that attitude importance determines whether attitudes act like categories (i.e., attitudes are only stable in discrete states of positive or negative evaluations) or dimensions (i.e., attitudes can take any place on the dimension ranging from positive to negative): Important attitudes act more like categories and unimportant attitudes act more like dimensions. Studies that focused on political attitudes provided support for this hypothesis (Latané & Nowak, 1994; Liu & Latané, 1998; van der Maas, Kolstein, & van der Pligt, 2003). It is our view that important attitudes or more generally, strong attitudes, act as categories because strong attitudes correspond to highly connected networks.

Linking attitude strength to the connectivity of attitude networks can also shed more light on the consequences of ambivalence. In relation to attitude strength, ambivalence is a puzzling phenomenon because, on the one hand, ambivalent attitudes are weak predictors of behavior (Armitage & Conner, 2000; Conner & Sparks, 2002), and on the other hand, have high impact on information processing (Jonas, Diehl, & Bromer, 1997; Nordgren, van Harreveld, & van der Pligt, 2006). High impact on information processing of ambivalent attitudes was mostly found in the context of felt ambivalence (i.e., ambivalence that causes discomfort; e.g. van Harreveld, van der Pligt, & de Liver, 2009; van Harreveld, Rutjens, Rotteveel, Nordgren, & van der Pligt, 2009). In our view, felt ambivalence is likely to arise in highly connected networks with unaligned evaluative reactions because highly connected networks are in an unstable state when the nodes are not aligned (Cramer et al., 2013). This proposition can explain why ambivalent attitudes are less predictive of behavior while they do influence information processing. Due to the high connectivity, motivation arises to process information in a way that resolves the unstable state of ambivalence (cf., Nordgren et al., 2006; van Harreveld, Nohlen, & Schneider, 2015; van Harreveld, van der Pligt, & de Liver, 2009) but because the evaluative reactions are not aligned yet, the impact on behavior is low.

To summarize, conceptualizing strong attitude as highly connected networks provides a framework in which the key features of attitude strength — stability, resistance and impact on behavior and information processing — as well as attributes related to attitude strength, such as importance, elaboration and extremity can be integrated. Furthermore, this conceptualization sheds more light on the underlying process of the catastrophe model of attitudes and can explain why ambivalent attitudes can be regarded as both weak and strong.

2.3 Discussion

In the present chapter, we introduced the Causal Attitude Network (CAN) model as a formalized measurement model of attitudes. In this model, attitudes are conceptualized as networks of interacting evaluative reactions (e.g., beliefs, feelings and behaviors toward an attitude object). Interactions arise through direct causal connections and mechanisms that support evaluative consistency. Attitude networks are driven by the trade-off between optimization (i.e., consistency between evaluative reactions) and accuracy. This trade-off results in a small-world structure, in which evaluative reactions, that are similar to each other, tend to cluster. Conceptualizing attitudes as networks provides testable hypotheses for attitude change (e.g., change in an evaluative reaction will foremost affect the cluster it belongs to) and a parsimonious explanation for the differences between strong and weak attitudes by conceptualizing attitude strength as connectivity of attitude networks. Initial empirical tests of the CAN model support both the proposed small-world structure of attitude networks and the hypothesis that strong attitudes correspond to highly connected attitude networks.

2.3.1 Extensions of the CAN Model

The CAN model as presented in this chapter can only deal with binary data but items are often assessed on nominal or continuous scales in the attitude literature. A natural extension of the model therefore would be to allow for nominal or continuous data. Such extensions are relatively straightforward as there are both Markov Random Field models available for nominal data (Potts model; Wu, 1982) and continuous data (Gaussian Random Field; Lauritzen, 1996) and their application to psychometric data is currently under development (Epskamp, Maris, et al., 2018; van Borkulo et al., 2014). In our view, whether evaluative reactions are binary, categorical or continuous variables is a challenging question for research in itself and we would urge attitude researchers to use models developed to test the underlying distribution of evaluative reactions (e.g., De Boeck, Wilson, & Acton, 2005). Connectivity of the attitude network might provide a tentative answer of whether evaluative reactions represent dimensions or categories. As we discussed in the section on attitude strength, highly connected attitude networks behave more like categories and weakly connected attitude networks behave more like dimensions. It is likely that the overall behavior of the attitude network also extends to the behavior of the evaluative reactions, so that evaluative reactions in a highly connected network behave more like categories and evaluative reactions in a weakly connected network behave more like dimensions.

Another possible extension of the CAN model would be to use time-series data to estimate individual attitude networks. The measurement model that we developed in this chapter is only suited to cross-sectional data but related approaches can be used to model time-series data (Bringmann et al., 2013). While estimating networks on cross-sectional data provides useful information on the general structure of the attitude networks in the population, modeling individual attitude networks on time-series data provides a powerful tool to detect individual differences in attitude structure. Several studies indicate that individuals differ in their attitude structure (e.g., Huskinson & Haddock, 2004; van der Pligt, de Vries, Manstead, & van Harreveld, 2000; van der Pligt et al., 2000; van

Harreveld, van der Pligt, de Vries, & Andreas, 2000) but statistical tools to model such differences have, as yet, been lacking. It is our view that network analysis can fill this gap.

A related possibility to extend the CAN model would be to model the sizes of the attitude networks of different individuals. It is likely that individuals will differ in how many evaluative reactions they have toward an attitude object and that therefore their attitude networks consist of different numbers of nodes. As we have argued here, the size of the network has implications for the dynamics of the network. To measure, however, how many evaluative reactions an individual has toward an attitude object new approaches to the assessment of attitudes have to be developed.

2.3.2 The Challenge to Assess Attitude Networks

Devising attitude questionnaires that are tailored to the theoretical background of the CAN model in some instances contrasts markedly with devising attitude questionnaires from a latent variable perspective. For example, developing questionnaires that can be used to asses virtually all attitudes is reasonable from a latent variable perspective (cf., Crites et al., 1994), as items in the questionnaires are simply indicators, but it is not necessarily reasonable from a network perspective. There is no a priori reason to assume that every attitude network consists of the same nodes. It is, for example, very likely that some emotions are often experienced toward some attitude objects (e.g., anger toward presidential candidates), while they are virtually never experienced toward other attitude objects (e.g., anger toward a detergent brand).

Another implication of the CAN model for the assessment of attitudes that differs from the implications of latent variable models is that researchers should strive to measure all relevant evaluative reactions because otherwise the danger arises that one measures not the whole attitude network but only parts of it. As evaluative reactions in the CAN model represent autonomous causal entities, omissions of relevant evaluative reactions would decrease the validity of the model (in contrast to only decreasing reliability in latent variable models).

To construct attitude questionnaires from a network perspective, a theory-driven approach to questionnaire construction instead of an empirically-driven approach should be adopted (see Borsboom, Mellenbergh, & van Heerden, 2004). A challenge in the construction of such attitude questionnaires becomes to assess all relevant evaluative reactions in the attitude network. This approach shifts the focus of attitude questionnaire construction from striving for internal consistency to striving for assessing attitudes comprehensively. As nodes in attitude networks represent autonomous entities, they might in some instances correlate only weakly — internal consistency and validity of the attitude questionnaire might therefore be incompatible in some instances. It is therefore our view that estimating reliability of the whole attitude questionnaire is of limited value — it is, however, possible to assess each evaluative reaction with several indicators. In this case, reliability can be estimated to investigate whether the different questions tapping one evaluative reaction are likely to indeed measure a single entity. One way to go about constructing comprehensive attitude questionnaires might be to assess, in an open-ended questionnaire, which evaluative reactions are most common for the attitude object of interest.

After having constructed a comprehensive attitude questionnaire, the next challenge is to assess whether individuals differ in the number of evaluative reactions they have. As we already discussed here, it is likely that individuals with more knowledge of the attitude object have larger attitude networks. Furthermore, elaboration is also likely to increase the number of evaluative reactions one has (cf., Tesser & Leone, 1977; W. Wood et al., 1995).

Assessing whether an evaluative reaction represents a relevant node in an individual's attitude network might be possible by measuring how long it takes a participant to respond to an item. An item that reflects a relevant node in an individual's attitude network should be judged faster than an item that does not, because subjectively important attitudinal beliefs are judged faster (van Harreveld et al., 2000) and accessible attitudes are more easily retrieved from memory (Fazio & Williams, 1986). Another possibility to distinguish between irrelevant and relevant nodes might be to assess the salience or importance of a given belief (e.g., Ajzen, 1991; van der Pligt et al., 2000). The more salient or important a given belief, the more likely it would be that this belief represents a relevant node in the attitude network. Using this technique to distinguish between relevant and irrelevant nodes would be relatively straightforward, as instruments to assess the importance of a given belief have already been developed (e.g., van der Pligt et al., 2000).

The distinction between irrelevant or nonexistent nodes and relevant nodes bears some resemblance to the distinction between attitudes and non-attitudes (Converse, 1970), which was further developed by Fazio, Sanbonmatsu, Powell, and Kardes (1986). They argued that the distinction between attitudes and non-attitudes should be seen as a continuum rather than a dichotomy. From a network perspective, attitude networks that consist of only few nodes that are only weakly connected lie more at the non-attitude end of the continuum and attitude networks that consist of several nodes that are strongly connected lie more at the attitude end of the continuum.

Another issue in the assessment of attitude networks concerns the boundary specification problem in network analysis (Laumann, Marsden, & Prensky, 1989). The boundary specification problem refers to the difficulty of deciding which entities form part of the network. In attitude networks, it is for example, difficult to decide whether the attitude toward a presidential candidate and the attitude toward the candidate's party represent two distinct networks or one large network. There are, however, algorithms available that detect so-called community structures in the data (e.g., Clauset, Newman, & Moore, 2004; Girvan, Newman, & Moore, 2002; Newman, 2004, 2006; Newman & Girvan, 2004). A community structure refers to nodes that are highly interconnected and that only share weak connections with other community structures. A possible solution to the former example might thus be to assess both evaluative reactions toward the presidential candidate and her party and then apply an algorithm to detect community structures in the data. Based on this, one could then investigate whether the two attitudes represent two distinct networks or one large network.

A problematic aspect of applying community algorithms to detect the boundaries of attitude networks is that researchers have to assess a large set of items if they want to assess each variable that might be closely connected to the attitude network. Especially, if researchers want to focus on individual attitude networks, the set of items can quickly become too large to assess them frequently. However, this issue is less problematic for cross-sectional research. It would be possible to administer an extensive questionnaire with the most important variables that might be related to the attitude network to a group of individuals. Assessing the boundaries of attitude networks therefore currently needs to be done at the group level instead of at the individual level.

2.3.3 Future Study of the CAN Model

We now turn to some possible avenues for future study of the CAN model. Related to the implications on assessment of attitude networks, a possibility for future study would be to model responses on implicit measurements of attitudes with the CAN model. The currently most prominent reaction-time based measure of attitudes, the Implicit Association Test (IAT; Greenwald et al., 1998), however, is not suited to investigate attitudes from a network perspective, as it is not possible to analyze single items in this measure. Adopted variants of the Affect Misattribution Procedure (AMP; B. K. Payne, Cheng, Govorun, & Stewart, 2005), on the other hand, can be used to measure specific evaluative reactions (stereotype misperception task; Krieglmeyer & Sherman, 2012; emotion misattribution procedure; Rohr, Degner, & Wentura, 2014). However, different theoretical viewpoints on responses toward implicit measures of attitudes have different implications for which nodes should be measured with implicit measures. Two influential models of implicit measures of attitudes are the Motivation and Opportunity as Determinants of behavior (MODE) model (Fazio, 1990; Fazio & Olson, 2003b) and the Associative-Propositional Evaluation (APE) model (Gawronski & Bodenhausen, 2006, 2007, 2011).

From the perspective of the MODE model, implicit measures of attitudes are regarded as more 'pure' estimates of attitudes because they limit the opportunity to conceal the attitude (Fazio & Olson, 2003b). A central postulate of the MODE model regarding the relation between implicit and explicit measures of attitudes holds that the different measures show high correspondence when the attitude object is a socially non-sensitive issue (e.g., attitudes toward presidential candidates) and low correspondence when the attitude object is a socially sensitive issue (e.g., prejudice). From the perspective of the MODE model, evaluative reactions toward sensitive issues should thus be measured using implicit measurements. Measuring evaluative reactions toward non-sensitive issues, on the other hand, would not require the use of implicit measurements.

From the perspective of the APE model, responses on implicit measures of attitudes represent affective 'gut' reactions and responses on explicit measures of attitudes represent deliberative judgments that are propositional in nature (Gawronski & Bodenhausen, 2006). What follows from this postulate is that implicit measures would be better suited to measure evaluative reactions that represent feelings toward an attitude object and explicit measures would be better suited to measure beliefs toward an attitude object. To acquire a complete picture of an attitude network, researchers thus would have to use combinations of implicit and explicit measures.

We emphasize that arguments against conceptualizing attitudes as latent variables also pertain to models focusing on attitudes assessed with implicit measures. In our view, implicit measures of attitudes do not acknowledge the inherent complexity of attitudes as these measures treat attitudes as unidimensional continua, which is clearly expressed by how attitudes are assessed with implicit measures: in the four most used implicit procedures to assess attitudes (Nosek, Hawkins, & Frazier, 2011) — IAT, AMP, Go/No-Go Association Task (Nosek & Banaji, 2001), and evaluative priming procedures (Fazio, Jackson, Dunton, & Williams, 1995) — attitudes are reduced to a single score, which presupposes them to be unidimensional. As we have shown in this chapter, however, a conceptualization of attitudes that instead treats attitudes as complex systems has many advantages over treating attitudes as unidimensional continua. Developing implicit measures of attitudes from a network perspective is thus likely to further the understanding of attitudes to a greater extent then imposing the (untested) assumption that attitudes are unidimensional.

As we argued in this chapter, the CAN model also aids the progress of integrating connectionist models of attitudes with empirical research on attitudes. We showed that the basic tenet of connectionist models of attitudes — attitudes are conceived as a product of a network of interrelated nodes — is also a realistic conceptualization of empirical data on attitudes. A next step in integrating connectionist models with empirical data would be to use empirically estimated attitude networks as input for data simulation (see van Borkulo, Borsboom, Nivard, & Cramer, 2011 for an example of how an empirically estimated network can be used for data simulation in Netlogo; Wilensky, 1999). The advantage of using empirically estimated attitude networks as input for data simulation is that the connection strengths between the network nodes are no longer based on arbitrary choices, but can be grounded in empirical data. In addition, while the CAN model can be used as a tool for data simulation, it can also produce empirical predictions on the dynamic behavior of estimated attitude networks.

We already discussed two central empirical predictions of the CAN model in the context of persuasion: (a) changes in a given evaluative reaction will foremost affect closely connected evaluative reactions, and (b) successful persuasion directed at highly central evaluative nodes is more likely to result in a fundamental change in the attitude network. A related prediction holds that evaluative reactions with high closeness have the highest impact on decisions. This prediction flows from the notion that evaluative reactions with high closeness are the most influential nodes in a network. While evaluative reactions with high closeness are not necessarily the nodes that are directly related to the decision, they can still have a profound influence on the decision through their influence on the other evaluative reactions in the network.

While persuasion research might benefit from investigating closeness of nodes, research on ambivalence might benefit from investigating the connectivity of attitude networks. The connectivity of attitude networks is particularly relevant to the distinction between potential and felt ambivalence (e.g., Newby-Clark et al., 2002). Potential ambivalence refers to ambivalent evaluations and felt ambivalence refers to psychological discomfort resulting from potential ambivalence. Some factors that influence whether potential ambivalence results in felt ambivalence have been identified. For example, simultaneous accessibility of ambivalent evaluations heightens discomfort resulting from ambivalence (Newby-Clark et al., 2002) and ambivalent attitudes that are relevant for decision-making are likely to cause aversive feelings (van Harreveld, van der Pligt, & de Liver, 2009; van Harreveld, Rutjens, et al., 2009). A general framework to integrate these different findings, however, has not yet been proposed. The CAN model might provide such a framework.

To see this, it is important to consider the notion of *attractor states*. Attractor states refer to states that a dynamical system is driven to (Alligood, Sauer, Yorke, & Crawford, 1997). Which attractor states exist in an attitude network likely depends on the connectivity of the network. Attitude networks that are highly connected probably display only two attractor states (positive or negative) because the strong connections between the nodes force all nodes to align to the same state. In a weakly connected network, several attractor states are likely to exist, as only some nodes have to settle in the same state.[6]

The drive of the attitude network to settle in an attractor state might reflect aversive feelings caused by ambivalence. Highly connected attitude networks would thus be more likely to cause aversive feelings in the light of ambivalence than weakly connected attitude networks. This is because highly connected networks must settle in one of only two attractor states, while weakly connected networks can settle in one of several attractor states. Factors that heighten the probability of felt ambivalence might thus be integrated by proposing that they cause stronger connectivity in attitude networks.

Another research area, in which network analysis and the CAN model in particular might provide new insights, is research on the development of attitudes over time. This area has received somewhat limited attention and the CAN model might aid progress in this area as network analysis is well suited to model developmental changes (Bringmann et al., 2013). A central postulate of the CAN model is that attitude networks

[6]This point is also illustrated in the section on attitude strength. As can be seen in Table 2.2, configurations of the attitude network, in which not all evaluative reactions are aligned, are more probable in weakly connected networks than in highly connected networks.

grow (i.e., new evaluative reactions attach to older evaluative reactions). This growth might depend on how often an individual elaborates on the attitude (Monroe & Read, 2008). As we already discussed here, the CAN model predicts that the structure of attitude growth will be driven by the similarity and popularity of evaluative reactions.

Another issue in research on attitudes that might benefit from investigating attitudes from the perspective of the CAN model is the question of to what extent attitudes are formed through environmental factors and genetic factors, and how these factors interact in the shaping of attitudes. While studies investigating influences on attitudes generally have focused on environmental influences (e.g., persuasion), genetic influences on attitudes have also been observed (Eaves, Eysenck, & Martin, 1989; Olson, Vernon, Harris, & Jang, 2001; Tesser, 1993).

Research on personality has shown that different genes influence different nodes in a network to a different extent (Cramer et al., 2012) and it is likely that a similar pattern exists for nodes in attitude networks. For example, fearful reactions to attitude objects might have a stronger genetic basis than other nodes in attitude networks because conservative attitudes were shown to be predicted by individual differences in arousal in response to threatening stimuli (Oxley et al., 2008), which in turn are predicted by amygdala activation (e.g., Larson et al., 2006), while amygdala activation was found to have a genetic basis (e.g., Hariri et al., 2002). A possible mechanism of how a conservative attitude network might develop is that an individual, who is genetically predisposed to fearful reactions, reacts fearfully to threats to social order and therefore is more vulnerable to persuasion attempts promoting that political decisions foremost must guarantee safety to social order. Network analysis would aid in identifying such processes.

2.3.4 Conclusion

In this chapter we introduced the CAN model and argued (a) that the Ising model provides a psychometrically realistic formalization of attitudes, (b) that attitude networks conform to a small-world structure and (c) network connectivity provides a mathematically-formalized conceptualization of attitude strength. The CAN model shows promise in integrating existing findings, which are hitherto disparate, and in furthering the development of new insights in several areas of attitude research. For these reasons, the CAN model is likely to make significant contributions to the integration of different areas of attitude research, and to further new insights into the complex concept of attitude.

Chapter 3

Network Analysis on Attitudes

Abstract

In this chapter, we provide a brief tutorial on the estimation, analysis, and simulation on attitude networks using the programming language R. We first discuss what a network is and subsequently show how one can estimate a regularized network on typical attitude data. For this we use open-access data on the attitudes toward Barack Obama during the 2012 American presidential election. Second, we show how one can calculate standard network measures, such as community structure, centrality, and connectivity on this estimated attitude network. Third, we show how one can simulate from an estimated attitude network to derive predictions from attitude networks. By this, we highlight that network theory provides a framework for both testing and developing formalized hypotheses on attitudes and related core social psychological constructs.

This chapter has been adapted from: Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2017). Network analysis on attitudes: A brief tutorial. *Social Psychological and Personality Science*, *8*, 528-527.

3.1 Introduction

Network theory might be the most interdisciplinary framework to date (e.g., Barabási, 2011; Borgatti et al., 2009; Wasserman & Faust, 1994; Watts & Strogatz, 1998). In the past decade, network analysis has become increasingly important in psychology, where it was introduced as a psychometric framework for the representation of clinical and cognitive psychological constructs (e.g., Borsboom, 2008; Cramer et al., 2010; Boschloo, van Borkulo, Borsboom, & Schoevers, 2016; McNally et al., 2015; van Borkulo et al., 2015; van de Leemput et al., 2014; van der Maas et al., 2006). More recently, network modeling has also been introduced into social psychology in the form of the Causal Attitude Network (CAN) model, which conceptualizes attitudes as networks of causally interacting evaluative reactions (i.e., beliefs, feelings, and behaviors toward an attitude object; Dalege et al., 2016).

The CAN model provides a promising theoretical framework for research on attitudes and other related social psychological core constructs, such as self-concept, prejudice, and interpersonal attraction. The generality of the data-analytic framework and simulation techniques that the CAN model uses, invites the application of the model to many different attitude objects. The model may be thus useful to researchers with various substantive interests and here we provide a brief tutorial on how to apply network theory to attitudes. While we focus on attitude networks, the techniques discussed here can easily be transferred to related constructs.

The outline of this chapter is as follows. First, we briefly discuss what a network is. We then illustrate network estimation from data on attitudes and calculation of standard network measures. Finally, we show how one can use simulations to investigate consequences of manipulating structural properties of an estimated network. The tutorial provides code in the programming language R (R Core Team, 2013) and we assume basic familiarity with R. Readers, who are not familiar with R, are referred to Torfs and Bauer (2014) for an excellent introduction to R.

3.1.1 What is a Network?

A network (called a graph in the mathematical literature) is an abstract representation of a system of entities or variables (i.e., nodes) that have some form of connection with each other (i.e., edges). Nodes can represent any entity or variable, like people, brain regions, or evaluative reactions, and edges can represent any form of connection, like friendship, associations in blood flow, or direct causal connections. Figure 3.1 shows an

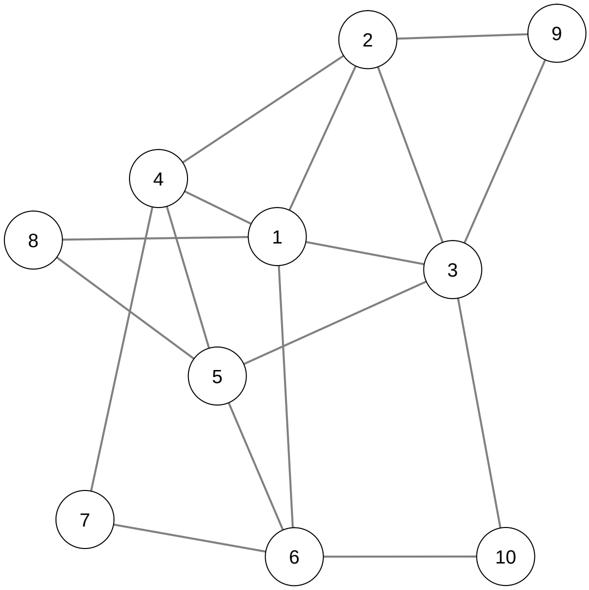


Figure 3.1: Example of a simple undirected and unweighted network with ten nodes (represented by circles) and 17 edges (represented by lines).

example of a simple network. Networks can be unweighted (i.e., edges are either present or absent) or weighted (i.e., edges can also differ in magnitude) and networks can be directed (i.e., edges indicate the direction of the connection) or undirected (i.e., edges show no direction). Here we focus on weighted undirected networks, because the techniques for estimating directed networks from correlational data require assumptions that are often not tenable in psychological data (e.g., there are no reciprocal influences between variables; Costantini et al., 2015). Excellent and thorough introductions to network theory are provided in Kolaczyk (2009) and Newman (2010).

From a network perspective, attitudes are systems of causally interacting evaluative reactions that strive for a coherent representation of the attitude object (Dalege et al., 2016). Based on this basic idea, Dalege et al. (2016) developed the CAN model that links research on attitudes to network theory. Important tenets of the CAN model are that attitude networks show a high degree of clustering, with similar evaluative reactions exerting stronger influence on each other than dissimilar evaluative reactions (e.g., judging a person as honest exerts a stronger influence on

Table 3.1: List of Items Tapping Evaluative Reactions and Their Abbreviations

Item	Abbreviation
Items tapping beliefs	
"is moral"	Mor
"would provide strong leadership"	Led
"really cares about people like you"	Car
"is knowledgeable?	Kno
"is intelligent?	Int
"is honest?	Hns
Items tapping feelings	
"angry?	Ang
"hopeful?	Hop
"afraid of him?	Afr
"proud?	Prd

Note. Participants rated whether the items tapping beliefs described Barack Obama and whether they ever felt the feelings described by the items tapping feelings toward Obama.

judging this person as caring and vice versa than judging this person as intelligent) and that strong attitudes correspond to highly connected attitude networks. The CAN model further assumes that the dynamics of attitude networks can be captured by the Ising (1925) model, which originated from statistical physics. Here, we use this assumption to simulate attitude networks based on the CAN model.

3.1.2 Network Analysis

In this section, we show how networks can be estimated from responses to attitude items and how one can calculate common network measures on these estimated networks. We focus on binary data, as the simulation based on the estimated networks that we discuss later is based on a model of binary variables. For estimating networks from continuous data or from data involving different types of variables, the interested reader is referred to Epskamp, Borsboom, and Fried (2018) and Haslbeck (2017), respectively, and to Appendix B. For the illustration of the techniques, we use the open-access data from the American National Election Study (ANES) of 2012 (available at www.electionstudies.org) on evaluative reactions toward Barack Obama (see Table 3.1 for an overview of the evaluative reactions). In the ANES of 2012, 5914 individuals, representative of the adult US American population, participated.

3.1.3 Estimation of Attitude Networks

For the estimation of attitude networks, we use the eLasso-procedure (van Borkulo et al., 2014). The eLasso-procedure regresses each variable on all other variables in turn and each regression function is subjected to regularization to reduce the size of the statistical problem of regressing a variable on a large number of variables and cope with the problem of multicollinearity in a data set involving a large number of variables (see Friedman et al., 2008; Tibshirani, 1996). The best-fitting regression function is selected using the Extended Bayesian Information Criterion, as described in Foygel and Drton (2010). The independent variables included in the selected regression function define the nodes that the dependent variable is connected to by edges, which are weighted by the regression parameters.

If the observed data are indeed realizations of a (sparse) network structure, this technique provides an accurate estimation of this network (van Borkulo et al., 2014). In this case, the resulting network can also be regarded as a causal skeleton, in which the edges represent putative causal associations that can be either directed or reciprocal.

If one does not want to make the assumption that a causal network underlies the data, a network may still be highly useful, because the edges represent how well the connected variables predict each other if all other observed variables are being held constant. Thus, networks can be used to gain insight into the causal structure of the data, but also to merely describe the pattern of predictive relations in a dataset, or to represent the correlation structure of the data (Epskamp et al., 2012). In some instances one might be willing to make the assumption that some of the variables measured form a causal system, while other variables might be worth controlling for without making the assumption that they form part of the causal system (i.e., covariates). We included a description in Appendix B how to deal with covariates in network estimation.

Simulation studies have shown that for binary data, sample sizes of 500 are generally sufficient to estimate networks of low and moderate size (i.e., networks consisting of 10-30 nodes) and that for large networks (i.e., networks consisting of 100 nodes) a sample size of 1000 is needed (van Borkulo et al., 2014). Regarding networks based on continuous data, sample sizes of 250 are generally sufficient for networks of moderate size (e.g., 25 nodes; Epskamp, 2016).

To estimate a network on the responses to the attitude items, which are stored in the data frame `Obama` (see the R code in Appendix C), we can use the function `IsingFit`, available in the R package *IsingFit* (van Borkulo &

Table 3.2: Weight Adjacency Matrix of the Obama Network

	Mor	Led	Car	Kno	Int	Hns	Ang	Hop	Afr	Prd
Mor	0	0.38	1.23	0.49	1.13	1.76	-0.22	0.19	-0.42	0.52
Led	0.38	0	0.8 1	.38	0.58	0.93	-0.84	0.33	-0.46	0.82
Car	1.23	0.8	0	0.65	0.68	1.38	-0.49	0.78	-0.48	0.86
Kno	0.49	1.38	0.65	0	2.66	0.67	0	0.56	-0.23	0.39
Int	1.13	0.58	0.68	2.66	0	0.86	0	0.31	0.24	0.25
Hns	1.76	0.93	1.38	0.67	0.86	0	-0.36	0.34	-0.73	0.42
Ang	-0.22	-0.84	-0.49	0	0	-0.36	0	-0.28	2.21	0
Hop	0.19	0.33	0.78	0.56	0.31	0.34	-0.28	0	-0.78	2.32
Afr	-0.42	-0.46	-0.48	-0.23	0.24	-0.73	2.21	-0.78	0	-0.16
Prd	0.52	0.82	0.86	0.39	0.25	0.42	0	2.32	-0.16	0

Note. Note: See Table 3.1 for the abbreviations of the nodes.

Epskamp, 2016) and save the results in the object ObamaFit:

```
ObamaFit <- IsingFit(Obama)
```

The object ObamaFit now contains the estimated network in the form of a weight adjacency matrix. A weight adjacency matrix has the same number of columns and rows as the number of nodes in the network and the values in the matrix represent the edge weights between the different nodes (see Appendix B for how to assess the stability of edge weights). The weight adjacency matrix of the Obama network can be called by using the command ObamaFit$weiadj, see Table 3.2. Each value above (below) the diagonal represents an edge from the node in the given row (column) to the node in the given column (row). As the Obama network is undirected, the weight adjacency matrix is symmetric. The values on the diagonal represent self-loops of the nodes and these are all 0 in the Obama network, as self-loops cannot be estimated using cross-sectional data. The object ObamaFit also contains the thresholds of the different nodes, which are the slopes of the regression equations of predicting a given node. The thresholds can be called with the command ObamaFit$thresholds.

To plot the network, we can use the function qgraph, available in the R package *qgraph* (Epskamp, Costantini, et al., 2018). The following command produces the network shown in Figure 3.2 (except for the color of the nodes; we come back to this issue in the section on community detection):

```
ObamaGraph <- qgraph(ObamaFit$weiadj, layout = 'spring', cut
    ↪ = .8)
```

The first argument in the function specifies the weight adjacency matrix to plot. The `layout` argument is used to plot the network using the Fruchterman-Reingold algorithm (Fruchterman & Reingold, 1991) that places strongly connected nodes close to each other. The `cut` argument specifies which edges should be plotted with higher width. All edges higher than the value specified in the `cut` argument are plotted with width according to their magnitude. Edges below the value specified in the `cut` argument only differ in color density. In the supplementary R code in Appendix C we also show how to plot a network using other color pallets.

3.1.4 Community Detection

Looking at the network shown in Figure 3.2, one can immediately recognize that nodes differ in their interconnectedness. To formalize this impression, we can use an algorithm that detects communities (or clusters) within the network. We use the walktrap algorithm (Pons & Latapy, 2005), as this algorithm performs well on psychological networks (Gates, Henry, Steinley, & Fair, 2016; Golino & Epskamp, 2017). An advantage of this algorithm compared to factor analyses is that it can detect dimensions of variables very well even when the different dimensions are highly correlated (Golino & Epskamp, 2017). To run the walktrap algorithm on the Obama network, we can use the function `cluster_walktrap`, available in the R package *igraph* (Amestoy et al., 2018). Before doing so, we have to create an *igraph* object containing the information on the Obama network and we have to take the absolute value of the edge weights, because the walktrap algorithm can only deal with positive edge values:

```
ObamaiGraph <- graph_from_adjacency_matrix(abs(ObamaFit$
   ↪ weiadj), 'undirected', weighted = TRUE, add.colnames =
   ↪  FALSE)
```

This command creates the *igraph* object `ObamaiGraph` containing the Obama network with absolute edge weights. The first argument specifies the input adjacency matrix, which is the weighted adjacency matrix of the Obama network with absolute values. The second and third arguments specify that we want an undirected and weighted network, respectively. The `add.colnames` argument can be ignored at the moment. We can now run the walktrap algorithm on the `ObamaiGraph` object:

```
ObamaCom <- cluster_walktrap(ObamaiGraph)
```

This function runs the walktrap algorithm on the Obama network and

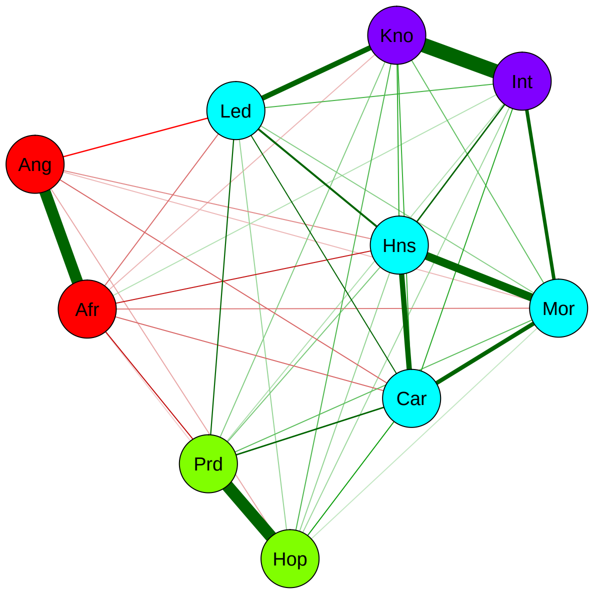


Figure 3.2: Network of the attitude toward Barack Obama. Nodes represent evaluative reactions and edges represent connections between evaluative reactions, with the edge width and color density corresponding to the strength of the connections. Green (red) edges represent positive (negative) connections. Colors of the nodes correspond to detected communities in the network. See Table 3.1 for the abbreviations of the nodes.

saves the results to the object `ObamaCom`. We can extract the found communities using the following command:

```
communities(ObamaCom)
```

To plot the network with the nodes being colored according to their community membership, we can use the following command, which creates Figure 3.2:

```
qgraph(ObamaFit$weiadj, layout = 'spring', cut = .8, groups =
    ↪  communities(ObamaCom), legend = FALSE)
```

The `groups` argument is used to assign nodes to groups and the communities function extracts a list containing the community membership of each node. Here, the `add.colnames = FALSE` argument of the `graph_from_adjacency_matrix` function comes into play, because this argument results in the nodes being numbered instead of having the variable names. This is needed, because the `groups` argument can only handle numeric values. The `legend = FALSE` argument is used because plotting the legend of the communities would not be informative.

The interpretation of the results of the community detection algorithm is straightforward: The red nodes represent negative feelings toward Barack Obama, the green nodes represent positive feelings toward Obama, the light blue nodes represent judgments pertaining mainly to interpersonal warmth (with judging Obama as a good leader also belonging to this community), and the purple nodes represent judgments pertaining to Obama's competence. The community structure of the Obama network is also in line with the postulate of the CAN model that similar evaluative reactions are likely to cluster (Dalege et al., 2016).

3.1.5 Node Centrality

While community detection can be used to inspect the global structure of the network, centrality of nodes can be used to inspect the structural importance of the different nodes. Centrality of nodes can, for example, be used to infer which evaluative reactions most likely influence decision-making (Dalege, Borsboom, van Harreveld, Waldorp, & van der Maas, 2017) and which evaluative reactions would be the most effective targets for persuasion attempts (see section on network simulation). The probably most popular measures of centrality are degree (called strength in weighted networks), closeness, and betweenness (Barrat, Barthélemy, Pastor-Satorras, & Vespignani, 2004; Freeman, 1978; Opsahl et al., 2010). The most straightforward of these indices is strength, as it is the sum of the absolute edge values connected to a given node and therefore represents the direct influence a given node has on the network.

Closeness and betweenness both depend on the concept of shortest path lengths. The shortest path length between two given nodes refers to the shortest distance between these two nodes based on the edges that directly or indirectly connect these two nodes. Dijkstra's algorithm is used to find shortest path lengths in weighted networks (Brandes, 2001; Dijkstra, 1959; Newman, 2001). Based on this algorithm, shortest path lengths represent the inverse of edge weights that have to be "travelled" on the shortest path (see Appendix B for an illustration of shortest path lengths). Closeness sums the shortest path lengths between a given node and all other nodes in the network and takes the inverse of the resulting value. Therefore, closeness represents how likely it is that information from a given node "travels" through the whole network either directly or indirectly. Betweenness represents how strongly a given node can disrupt information flow in the network, as betweenness calculates the number of shortest paths a given node lies on.

To calculate the different centrality estimates for the Obama network, we can use the function centralityTable and to plot these indices, we can use the function centralityPlot, both available in *qgraph* (see Appendix B for how to assess the stability of centrality indices):

```
ObamaCen <- centralityTable(ObamaGraph, standardized = FALSE)
centralityPlot(ObamaGraph, scale = 'raw')
```

The first argument in these functions specifies the network for which we want to calculate the centrality indices and the standardized = FALSE and scale = 'raw' arguments specify that we want unstandardized centrality estimates. The plots produced by the centralityPlot function are shown in Figure 3.3. As can be seen, two nodes seem to have high structural importance: The node Led has the highest betweenness and the highest closeness, while the node Hns has the highest strength. Looking back at Figure 3.2, we can see that the reason for the node Led having the highest betweenness is probably that, while it belongs to the warmth community, it is also relatively closely connected to the two feeling communities and to the competence community. This node thus connects the different communities and changes in the different communities probably only affect the other communities when the Led node also changes. That the node Led is closely connected to all communities also explains why it is the node with the highest closeness. Change in this node is thus likely to affect large parts of the network. Hns has the highest strength because it has strong connections to the nodes Mrl, Car, Int, and Led. Change in the node Hns would thus strongly affect many other nodes.

3.1.6 Network Connectivity

While centrality of nodes provides information on how change in a given node would affect the rest of the network, network connectivity provides general information on the dynamics of the network, as network connectivity and dynamics are closely connected (e.g., Kolaczyk, 2009; Manrubia & Mikhailov, 1999; Scheffer et al., 2012; Watts, 2002).

A common index of network connectivity is the Average Shortest Path Length (L; West, 1996), which is the average of all shortest path lengths between all nodes in the network. A low L indicates high connectivity. To calculate the L of the Obama network, we first have to calculate the matrix, containing all shortest path lengths. This can be done using the function centrality, available in the R package *qgraph*. Then we need to

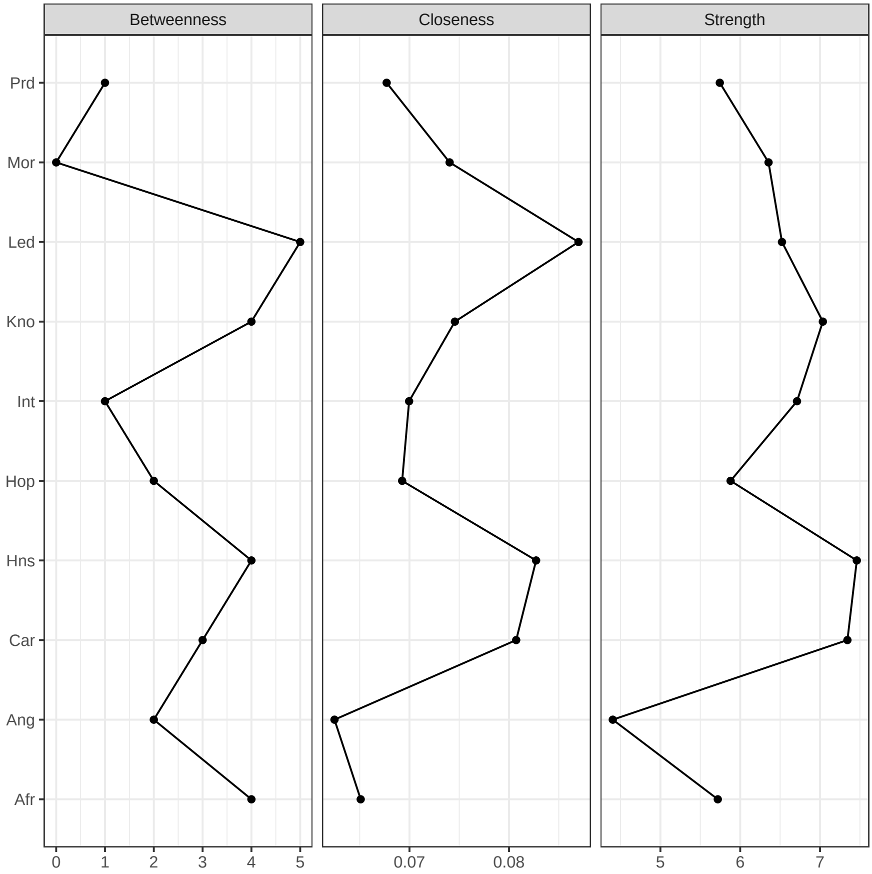


Figure 3.3: Centrality plot of the Obama network. Left (middle) [right] panel shows the betweenness (closeness) [strength] estimates for each node of the Obama network. See Table 3.1 for the abbreviations of the nodes.

select the upper triangle of this matrix and take the mean of the resulting values:

```
ObamaSPL <- centrality(ObamaGraph)$ShortestPathLengths
ObamaSPL <- ObamaSPL[upper.tri(ObamaSPL)]
ObamaASPL <- mean(ObamaSPL)
```

The object `ObamaASPL` now contains the L of the Obama network, which is equal to 1.53. In the section on network simulation, we illustrate some of the differences between highly and weakly connected networks.

3.1.7 Comparison of Networks

While it can be informative to study one attitude network, in many instances we want to compare networks of different attitudes. To do this, we can use the recently developed Network Comparison Test (NCT; van Borkulo, 2016). The NCT utilizes permutations (i.e., rearranging of samples) to test whether two networks are invariant with respect to global strength (i.e., the sum of all edge weights), network structure, and specific edge values. We illustrate the use of the NCT by comparing the network of the attitude toward Barack Obama to the network of the attitude toward Mitt Romney. To do so, we created two matched data frames on the evaluative reactions toward Barack Obama (ObamaComp) and Mitt Romney (RomneyComp), respectively (see the supplementary R code in Appendix C). To run the NCT on these data frames, we can use the function NCT, available in the R package *NetworkComparisonTest* (van Borkulo, 2016):

```
NCTObaRom <- NCT(ObamaComp, RomneyComp, it = 1000, binary.
   ↪ data = TRUE, paired = TRUE, test.edges = TRUE, edges =
   ↪  'all')
```

The first two arguments specify the data frames, we want to use to compare networks, the `it` argument specifies how many permutations will be performed, the `binary.data` argument specifies that we use binary data, the `paired` argument specifies that the two data frames contain responses of the same group of individuals, the `test.edges` argument specifies that we want to also test invariance of edges, and the `edges` argument specifies that we want to test the invariance of all edges in the network.

The NCT indicated that the global strengths of the networks did not differ significantly, but that the structures of the networks and some specific edges differed significantly. The difference in global strength was 1.12, p = .111. These values can be called using the commands `NCTObaRom$glstrinv.real` and `NCTObaRom$glstrinv.pval`. The structure of the networks was not invariant, as the maximum difference in edge weights of 0.72 was significant, p = .021. These values can be called using the commands `NCTObaRom$nwinv.real` and `NCTObaRom$nwinv.pval`. To investigate whether specific edges differed, we can use the command `NCTObaRom$einv.pvals`, which gives us the Bonferroni corrected p values for each edge. As can be seen in Figure 3.4, four edge weights differed significantly between the two networks, three of which are connected to the node Led. From this, we can conclude that the node Led has a different role in the Romney network than in the Obama network. In the Romney network, the node Led is more strongly (positively) connected to

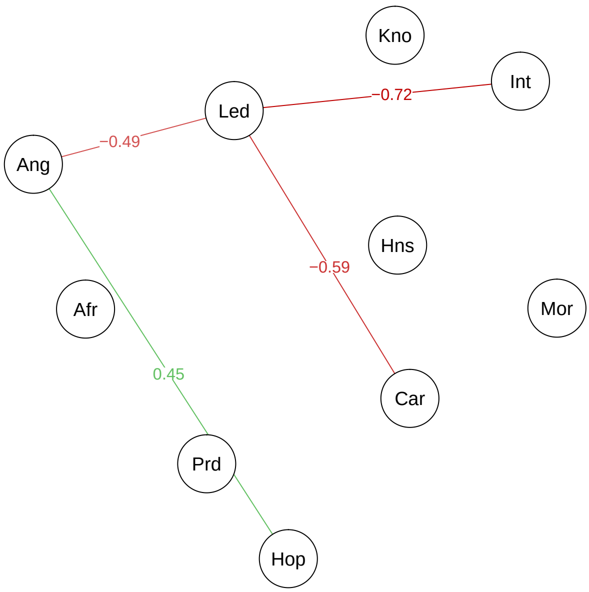


Figure 3.4: Edges that differ significantly between the Obama and the Romney network. Red (green) edges indicate edges that had a higher value in the Obama (Romney) network. Values indicate the difference between the edges in the Obama network and the Romney network.

other beliefs, while in the Obama network it is more strongly (negatively) connected to the feeling of anger.

3.2 Network Simulation

While *network analysis* can be used to describe systems, such as attitudes, *network simulation* can be used to make inferences on the dynamics of the system. Simulation of attitude networks can therefore help researchers to derive concrete hypotheses that in many instances cannot be derived by only studying the descriptives of the network. In this section, we first show how to simulate networks with varying connectivity and then show, as an illustration, how to simulate the results of change in nodes of varying centrality. This is helpful in determining at which nodes a persuasion attempt should be targeted.

To model dynamics of attitude networks, we make use of the Ising (1925) model, which represents an idealized model of probabilities that nodes in the network will be ”on” or ”off” (e.g., whether evaluative reactions will be endorsed or not). The Ising model consists of three classes of parameters. The first class represents continuous edge weights between nodes in the network that can be either positive, negative, or 0. Positive (negative) edge weights make it more likely that the connected nodes assume the same (different) state and the higher the magnitude of the weights, the more likely it is that nodes assume the same (different) state. The second class represents continuous thresholds of nodes that also can be either positive, negative, or 0. A positive (negative) threshold of a given node indicates that the node has the disposition to be ”on” (”off”) and the higher the magnitude, the stronger the disposition of the given node. Finally, the Ising model utilizes a temperature parameter that scales the entropy of the network model (Epskamp, Maris, et al., 2018; Wainwright & Jordan, 2008). High (low) temperature makes the network behave more (less) randomly by decreasing (increasing) the influence of both edge weights and thresholds. In attitude networks, temperature can be seen as a formal conceptualization of consistency pressures (Dalege, Borsboom, van Harreveld, Waldorp, & van der Maas, 2017).

3.2.1 Network Connectivity

To simulate dynamics of estimated networks, we can use the function `IsingSampler`, available in the R package *IsingSampler* (Epskamp, 2015). This function requires the network from which we want to simulate as input (i.e., the edge weights as well as the thresholds of nodes). We can extract this information from the object `ObamaFit` and we saved the relevant information in the object `SimInput` (see the supplementary R code in Appendix C). Note that we rescaled the negative evaluative reactions Ang and Afr, so that a positive (negative) score on all evaluative reactions indicates a positive (negative) evaluation.

To illustrate the consequences of varying network connectivity, we set all thresholds to 0 (implying all nodes have no disposition to be in a given state) and we manipulate the temperature of the network model. To simulate cases based on the Obama network with three different temperatures, we can use the following commands:

```
sampleLowTemp <- IsingSampler(1000, SimInput$graph, rep(0,10)
    ↪ , .4, responses = c(-1L,1L))
sampleMidTemp <- IsingSampler(1000, SimInput$graph, rep(0,10)
    ↪ , .8, responses = c(-1L,1L))
sampleHighTemp <- IsingSampler (1000, SimInput$graph, rep
    ↪ (0,10), 1.2, responses = c(-1L,1L))
```

The first argument of the `IsingSampler` function specifies the number of cases we want to simulate, the second argument specifies the network from which we want to simulate, the third argument specifies the thresholds, and the fourth argument specifies the inverse temperature. The `responses` argument is used to indicate that the network and thresholds we are simulating from are based on -1 and +1 responses. As can be seen in Figure 3.5, the connectivity of a network has fundamental implications for the distributions of the sum scores. The sum score is a useful measure of the overall state of the attitude and represents a measure of the global attitude toward Barack Obama in the current example. While the sum scores of a weakly connected network are normally distributed, sum scores of a highly connected network are distributed bimodally. This illustrates that psychological constructs, which are based on networks, can be either regarded as dimensions or categories depending on the connectivity of the network (cf., Borsboom et al., 2016). In the case of attitudes, this would mean that weakly connected attitude networks behave as dimensions (i.e., attitudes can take any evaluation ranging from negative to positive), while highly connected attitude networks behave as categories (i.e., attitudes are generally either positive or negative). For a thorough discussion of the implications of network connectivity for attitudes and especially attitude strength, the interested reader is referred to Dalege et al. (2018b).

3.2.2 Node Centrality

To illustrate the consequences of influencing nodes with different centrality, we show how to simulate targeting nodes differing in strength. As can be seen in Figure 3, the node with the highest (lowest) strength is the node Hns (Ang). Let us assume we want individuals to have a more positive attitude toward Obama and we focus on individuals, who have a slightly negative attitude toward Obama. We can simulate a group of individuals with moderately negative attitudes by setting all thresholds of the evaluative reactions to a moderately negative value (e.g., -.1):

```
SampleNeg <- IsingSampler(1000, SimInput$graph, rep(-
.1,10), responses = c(-1L,1L))
```

The mean sum score of this sample is -5.63 with a standard deviation of 6.54. To simulate a strong persuasion attempt focusing on the node Hns (Ang), we can set the node's threshold to 1. This represents a persuasion attempt to make individuals judge Obama as more honest (feel less angry toward Obama):

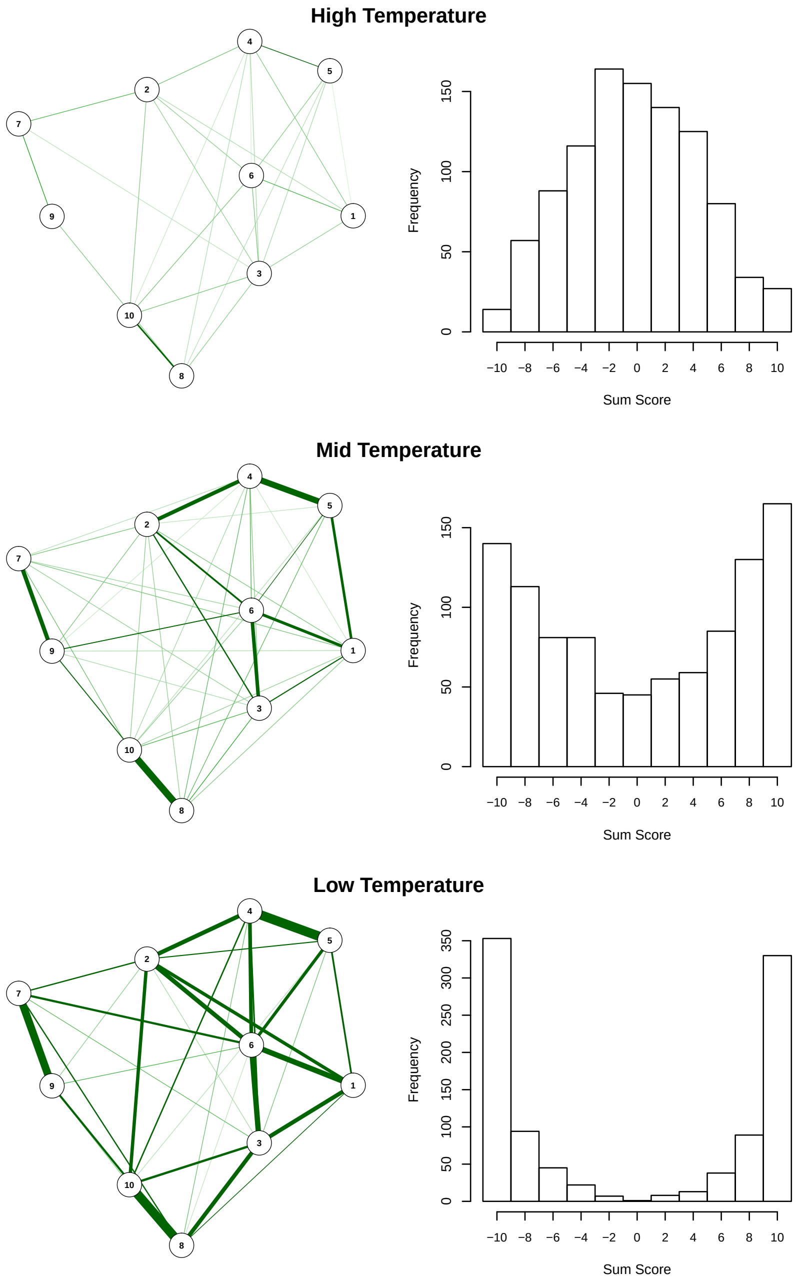


Figure 3.5: Networks with different temperatures based on the Obama network and their associated distributions of sum scores.

```
SampleHns <- IsingSampler(1000, SimInput$graph, c(rep(-.1,5),
   ↪  1, rep(-.1,4)), responses = c(-1L,1L))
SampleAng <- IsingSampler (1000, SimInput$graph, c(rep(-.1,6)
   ↪ , 1, rep(-.1,3)), responses = c(-1L,1L))
```

Important to note here is that the persuasion attempt in both situations was equally strong. Yet, the persuasion attempts significantly differed in their effectiveness reflected by the sum scores, $t(1998) = 6.27$, $p < .001$, $95\% CI[1.53, 2.91]$, $d = 0.28$. The persuasion attempt on the central node Hns resulted in a more positive attitude ($M = 1.18$, $SD = 7.99$) than the persuasion attempt on the peripheral node Ang ($M = -1.04$, $SD = 7.84$). From this simulation, we can thus derive the hypothesis that persuasion targeted at a node with high strength would result in higher attitude change than persuasion targeted at a node with low strength.

3.2.3 Testing Predictions from Simulations

In a recent study, Dalege, Borsboom, van Harreveld, Waldorp, and van der Maas (2017) used the Ising model to derive predictions regarding network structure and prediction of behavior. Simulations showed that highly connected attitude networks are more predictive of behavior than weakly connected attitude networks. To test this prediction, Dalege, Borsboom, van Harreveld, Waldorp, and van der Maas (2017) estimated attitude networks based on non-behavioral evaluative reactions toward presidential candidates (like the network discussed in the current chapter) and correlated the network connectivity with how well the attitude (based on the sum score of the evaluative reactions) predicted the voting decision. They found an almost perfect correlation between network connectivity and predictability of voting decisions, supporting the hypothesis derived from simulations on the Ising model. This study provides an illustration of how network simulation can aid hypothesis generation and also how to devise research on behavior-prediction in the attitude network framework.

3.3 Discussion

In this chapter, we provided a brief tutorial on estimating, analyzing, and simulating attitude networks. The combination of network analysis and simulation illustrates that network theory provides a framework to both test and develop formalized hypotheses on attitudes. It is our view that this makes network theory a unique framework, as it bridges the current gap between social psychological theorizing and psychometric theory.

Network analysis is a novel and still developing field in psychology. An important issue that needs to be tackled in the future is to devise strategies to make sure that all relevant nodes in a (attitude) network are measured. While currently the focus on questionnaire construction lies on selecting items that are highly interrelated (resulting in high reliability of the questionnaire), the network perspective on psychological constructs implies that the most important issue in questionnaire construction is to select items that provide a comprehensive picture of the measured construct. Such items might not be highly interrelated, which implies that validity and reliability of a questionnaire might in some instances be incompatible (Dalege et al., 2016).

With this tutorial, we aimed to provide an accessible introduction to network analysis and simulation on attitudes. While we have focused on attitudes in this tutorial, the techniques outlined here can also be applied to several other core social psychological constructs, such as self-concept, prejudice, and interpersonal attraction. It is therefore our view that network theory provides a promising framework to move our field forward.

Part II

Empirical Tests of the Causal Attitude Network (CAN) Model

Chapter 4

A Network Perspective on Attitude Strength: Testing the Connectivity Hypothesis

Abstract

Attitude strength is a key characteristic of attitudes. Strong attitudes are durable and impactful, while weak attitudes are fluctuating and inconsequential. Recently, the Causal Attitude Network (CAN) model was proposed as a comprehensive measurement model of attitudes, which conceptualizes attitudes as networks of causally connected evaluative reactions (i.e., beliefs, feelings, and behavior toward an attitude object). Here, we test the central postulate of the CAN model that highly connected attitude networks correspond to strong attitudes. We use data from the American National Election Studies 1980-2012 on attitudes toward presidential candidates (total n = 18,795). We first show that political interest predicts connectivity of attitude networks toward presidential candidates. Second, we show that connectivity is strongly related to two defining features of strong attitudes – stability of the attitude and the attitude's impact on behavior. We conclude that network theory provides a promising framework to advance the understanding of attitude strength.

This chapter has been adapted from: Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2018). A network perspective on attitude strength: Testing the connectivity hypothesis. *Social Psychological and Personality Science*. Advance online publication.

4.1 Introduction

While some attitudes are durable and impactful, other attitudes are largely inconsequential and fluctuating – in short, attitudes differ in their strength. These fundamental differences between strong and weak attitudes have spurred decennia of research in social psychology and several attributes related to attitude strength have been identified (e.g., importance, certainty, accessibility; for overviews see Cunningham & Luttrell, 2015; Krosnick & Petty, 1995; Visser et al., 2006). Recently, the Causal Attitude Network (CAN) model, which conceptualizes attitudes as networks of causally connected evaluative reactions (i.e., beliefs, feelings, and behaviors toward an attitude object; Dalege et al., 2016), has been proposed as a comprehensive measurement model of attitude. The CAN model holds that connectivity of attitude networks represents a formalized concetptualization of attitude strength. Here, we provide a first test of the connectivity hypothesis, which holds that highly connected attitude networks correspond to strong attitudes.

4.2 The Causal Attitude Network (CAN) Model

The basic premise of the CAN model is that an attitude is a system of evaluative reactions that influence each other (Dalege et al., 2016). Network modeling offers a natural representation of this hypothesis, because in network models, complex systems are modeled as a set of autonomous entities (i.e., nodes) and connections between these nodes (i.e., edges). Together, the set of nodes and edges defines the network structure (Newman, 2010). In attitude networks, nodes refer to evaluative reactions such as judging a presidential candidate as competent, charismatic and honest; feeling hope and pride about a presidential candidate; and showing support and voting for a presidential candidate. Edges in attitude networks represent bidirectional pairwise interactions between evaluative reactions (e.g., feeling hope about a presidential candidate may result from judging the candidate as being honest and vice versa). The CAN model is based on recent empirically derived network models as applied to psychopathology, personality, and cognitive psychology (e.g., Cramer et al., 2012, 2010; van der Maas et al., 2006) and on theoretical parrelel constraint satifaction models of attitude, which assume that the drive for cognitive consistency is the main force of attitude formation and change (e.g., Monroe & Read, 2008; Read, Vanman, & Miller, 1997; Shultz & Lepper, 1996; Simon, Snow, & Read, 2004; Simon, Stenstrom, & Read, 2015). The basic premise of the

CAN model – that attitudes are systems of interconnected beliefs, feelings and behaviors – is also reminiscent of work by McGuire (1990) on thought systems.

In the CAN model, the average level of the weights depends on how often an individual interacts with the attitude object in one way or another (e.g., thinking about or perceiving the attitude object). Based on the mechanism of constraint satisfaction that was implemented in a recent connectionist model of attitudes (Monroe & Read, 2008), the CAN model proposes that connections between evaluative reactions self-organize when the individual interacts with the attitude object. Attributes related to attitude strength, such as importance (Boninger et al., 1995) and elaboration (Petty et al., 1995), make it likely that an individual interacts with the attitude object and that the global connectivity of his or her attitude network increases as a result of this.

4.3 Global Connectivity of Attitude Networks

The level of global connectivity is a primary aspect of networks. Global connectivity in networks depends both on the number of connections between nodes and on the magnitude of the connection weights. In highly connected networks, changes in one node have stronger effects on other nodes. Thus, in highly connected attitude networks, evaluative reactions have more causal impact on each other than in weakly connected attitude networks. As a result, nodes in highly connected networks typically align their states, while nodes in weakly connected networks can vary relatively independently and show more random patterns (e.g., Kindermann & Snell, 1980; Scheffer et al., 2012). Next we will illustrate what this means in the context of attitudes, after which we will discuss why connectivity can be regarded as a formalized conceptualization of attitude strength.

Consider Alice and Bob, who both have a positive attitude toward Donald Trump. Suppose that both Alice's and Bob's evaluative reactions include judging Donald Trump as charismatic, honest and competent. Their attitudes, however, differ in their connectivity – Alice's network is highly connected while Bob's network is weakly connected (see Figure 4.1 for a graphic representation). As a result, if Alice were to receive information that is incongruent with the current state of a given evaluative reaction – say, she learns of a mistake Donald Trump made, implying that Donald Trump might not be as competent as Alice initially thought – strong causal connections between the judgments would create an un-

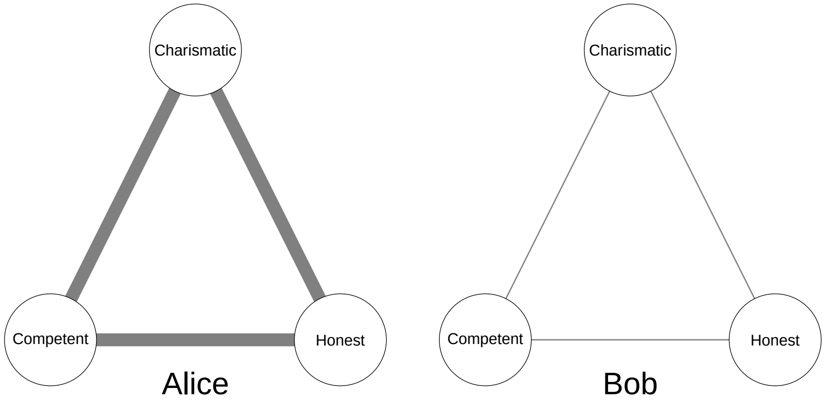


Figure 4.1: Representation of evaluative reactions that are part of a highly connected (Alice) or weakly connected (Bob) attitude network. Thickness of edges represents strength of (reciprocal) causal connections between the evaluative reactions.

stable state of the network (van Borkulo et al., 2014). To regain stability, Alice should either discard the information as unreliable or her other judgments should change, too – the former course of action, however, becomes more likely with increasing connectivity between the judgments up to the point where it becomes extremely difficult to change one of the evaluative reactions in isolation. If Bob, on the other hand, were to change his judgment that Donald Trump is competent, this change would not have much impact on his other judgments, because his network is weakly connected. Thus, conflicting evaluative reactions would disturb his attitude network to a much lesser extent, making it easier to change individual evaluative reactions. Highly connected attitudes thus result in consistent attitudes – an attribute that is related to attitude strength (Chaiken et al., 1995; Eagly & Chaiken, 1995; Judd & Krosnick, 1989; Judd, Krosnick, & Milburn, 1981). Network connectivity therefore provides a mechanistic explanation of why attitudes differ in their consistency (Dalege et al., 2016).

4.4 Network Connectivity and Attitude Strength

Dynamical properties of networks, such as their evolution in time and resistance to change, are intimately intertwined with network structure (e.g., Cramer et al., 2016; Kolaczyk, 2009; Manrubia & Mikhailov, 1999; Scheffer et al., 2012; Watts, 2002). It turns out that the dynamics of highly and weakly connected *networks* bear a striking resemblance to the defining features of strong and weak *attitudes* (Dalege et al., 2016). Note that this implies that connectivity of attitude networks predicts attitude strength. However, the CAN model does not exclude the possibility that also other factors independent of network connectivity predict attitude strength.

First of all, strong attitudes are durable – they are resistant to change and stable (e.g., Bassili, 1996; Bizer & Petty, 2005; Visser & Krosnick, 1998). Highly connected networks are more resistant to change than weakly connected networks. Weakly connected networks change roughly proportional to the external force put on the network's nodes (e.g., a persuasive message regarding Trump's incompetence), while a disproportionate amount of force is needed to instigate change in highly connected networks (Cramer et al., 2016). Furthermore, individual nodes in weakly connected networks will intrinsically show more random variation than nodes in highly connected networks because their behavior is to a larger extent controlled by random perturbations from outside the network (Kindermann & Snell, 1980; van Borkulo et al., 2014).

Second, strong attitudes are better predictors of behavior than weak attitudes (e.g., Bizer, Larsen, & Petty, 2011; Fazio & Williams, 1986; Holland et al., 2002). Because the CAN model treats behavior as part of the attitude network, factors that increase connectivity between non-behavioral evaluative reactions (i.e., beliefs and feelings) are expected to also influence the connectivity between non-behavioral evaluative reactions and behavioral evaluative reactions (e.g., behavioral decisions; Dalege, Borsboom, van Harreveld, Waldorp, & van der Maas, 2017). This implies that highly connected attitude networks should be more predictive of behavior.

Third, strong attitudes exert more influence on information processing – they direct attention and influence the way in which incoming information is integrated (e.g., Fazio & Williams, 1986; Houston & Fazio, 1989; Roskos-Ewoldsen & Fazio, 1992). While connectivity of attitude networks directly causes the former three features of attitude strength, the influence on information processing in our framework is probably indirect. As discussed above, successful persuasion is more likely to instigate conflict in highly connected networks. It is likely that individuals are motivated to avoid such conflict and are therefore motivated to integrate information in a way that does not disrupt the equilibrium of their attitude network. This motivation also might lead to heightened attention to attitude objects to early detect "attacks" on the attitude network.

Linking network connectivity to attitude strength thus might provide a formalized explanation how the effects of strong attitudes are manifested. However, empirical support for the hypothesis that network connectivity predicts attitude strength is still lacking. The aim of this chapter is to provide a first test of this hypothesis and thereby providing more empirical grounding of the CAN model.

4.5 Overview of the Present Research

To test the connectivity hypothesis, we applied network analysis to the open-access data of the American National Election Studies (ANES) between 1980–2012. These data sets involve large and representative samples and focused on both theoretically and practically highly relevant attitudes – attitudes toward presidential candidates. Furthermore, the ANES were used in several earlier lines of research into attitude strength and related issues (e.g., Alwin & Krosnick, 1991; Bizer et al., 2004; Crano, 1997; Eaton, Visser, Krosnick, & Anand, 2009; Judd et al., 1981; Krosnick, 1988; Krosnick & Alwin, 1989; Lusk & Judd, 1988; Visser, Krosnick, & Simmons, 2003).

To test the connectivity hypothesis, we first had to identify a variable likely to predict network connectivity. The CAN model assumes that connectivity of attitude networks depends on the amount of interaction with the attitude object and we therefore selected a measure of political interest as the predictor of network connectivity of attitudes toward presidential candidates. Our first analysis tested whether indeed political interest predicts attitudinal network connectivity.

While interest is related to attitude strength (e.g., Krosnick et al., 1993), it is not a defining feature of attitude strength. In the second analysis we therefore tested whether network connectivity predicts two defining features of attitude strength. First, we tested whether network connectivity predicts durability of attitudes by assessing attitude stability. Second, we tested whether network connectivity predicts impact of attitudes by assessing the impact of attitudes on behavior. The connectivity hypothesis holds that network connectivity positively predicts attitude strength.

4.6 Method

4.6.1 Participants

The ANES samples are large samples that are representative of the US American adult population. Questionnaires were administered in two surveys before and after each American presidential election from 1980 to

Table 4.1: Numbers of Participants Assigned to the Interest Groups for Each Election and Numbers of Participants who Had Missing Values on the Interest Variable

Election	Low interest	Intermediate interest	High interest	Missing values
1980	407 (121, 149)	692 (78, 100)	407 (50, 16)	49
1984	558 (132, 123)	1054 (123, 170)	638 (53, 81)	7
1988	509 (139, 134)	960 (102, 133)	567 (50, 57)	4
1992	419 (89, 60)	635 (101, 64)	304 (48, 20)	1
1996	399 (38, 87)	848 (31, 100)	467 (22, 33)	0
2000	396 (137, 111)	886 (148, 131)	525 (76, 44)	0
2004	186 (71, 31)	528 (102, 32)	498 (57, 20)	0
2008	378 (72, 87)	815 (89, 93)	1128 (88, 97)	1
2012	895 (73, 121)	2460 (68, 153)	2554 (43, 81)	5

Note. A subsample of the ANES 1996 already participated in the ANES 1992. The number of participants who completed the pre-election survey of the ANES 1996 and also participated in the ANES 1992 is 1,316. Numbers of excluded participants are shown in parentheses for the attitudes toward Democratic and Republican candidates, respectively.

2012 by the Center for Political Studies of the University of Michigan. The samples to which all relevant items were administered in the pre-election survey consisted of 18,795 adult participants in total (see Table 4.1 for number of participants per election). This large sample size provides us with high statistical power.

4.6.2 Measures

Relevant measures included non-behavioral evaluative reactions toward presidential candidates, a measurement of the interest participants had in the campaign, a measure of the global attitude toward presidential candidates assessed before and after each election, and an assessment of whom the participants voted for.

Evaluative reactions. Six to 16 items tapping beliefs regarding the presidential candidates and between four to eight items tapping feelings toward the presidential candidates were assessed in the pre-election surveys of the different studies. Note that some items were administered in every ANES or in most ANES, while other items were only administered infrequently, see Table 4.2.

For items tapping beliefs, participants were asked: ”In your opinion, does the phrase ’he...’ describe *the candidate*...?” and Table 2 shows the phrases that completed the items. In the ANES from 1980 to 2004, and for a subsample of the ANES of 2008 (N = 1133), items were assessed on a 4-point scale, with answer options 4 = ”extremely well”, 3 = ”quite well”,

Table 4.2: Included and excluded evaluative reactions.

Evaluative Reaction	Included (Excluded)
Items tapping beliefs	
"really cares about people like you"	1984–2012 (1984^D)
"is compassionate"	1984–1996* (1984^D, 1988)
"is decent"	1984, 1988 (1988^D)
"is dishonest"	1980, 2000, 2004
"would solve our economic problems"	1980
"gets things done"	1992, 1996 (1992^D)
"sets a good example"	1984 (1984^D)
"is fair"	1984 (1984^D)
"would develop good relations with other countries"	1980
"is hard-working"	1984 (1984^D)
"is honest"	1988–1996, 2008, 2012 (1988^R, 1992^D)
"is in touch with ordinary people"	1984 (1984^D)
"is inspiring"	1980, 1984, 1992, 1996 (1984^D)
"is intelligent"	1984–2012*
"can't make up his mind"	2004
"is kind"	1984 (1984^D)
"is knowledgeable"	1980–2012 (1988^D)
"would provide strong leadership"	1980–2012 (1984^D, 1992^D)
"is moral"	1980–2012 (1984^D, 1988, 2000^R)
"is optimistic"	2008
"is out of touch with ordinary people"	2000
"is religious"	1984 (1984)
"commands respect"	1984 (1984^D)
"understands people like you"	1984 (1984^D
"is power-hungry"	1980
"is weak"	1980 (1980^R, 2000^R)
Items tapping feelings	
"angry"	1980–2012 (1980^R)
"afraid of him"	1980–2012 (1980^R
"disgusted"	1980, 1984
"hopeful"	1980–2012
"proud"	1980–2012
"sympathetic toward him"	1980, 1984
"uneasy"	1980, 1984

Note. *In 1992, these items were only assessed for Bill Clinton, ^{D} the item was only excluded for the Democratic candidate, ^{R} the item was only excluded for the Republican candidate.

2 = "not too well", 1 = "not well at all". In order to use current network estimation software (van Borkulo et al., 2014), we dichotomized the data into two categories, one consisting of responses 1 and 2, and one consisting of responses 3 and 4. For a subsample of participants of the ANES of 2008 (N = 1133), and for all participants in the ANES of 2012, the items were assessed on a 5-point scale, with the answer options 5 = "extremely well", 4 = "quite well", 3 = "moderately well", 2 = "slightly well", 1 = "not well at all". Here, we assigned option 1, 2 and 3 in one category and option 4 and 5 in the other category.[1]

For items tapping feelings, participants were asked: "Has *the candidate* – because of the kind of person he is or because of something he has done, ever made you feel: ...?" and the feelings that completed the items are shown in Table 4.2. These items were assessed dichotomously, with the answer options 1 = "yes" and 0 = "no".

Political interest. Political interest was assessed by the item: "Some people don't pay much attention to political campaigns. How about you? Would you say that you have been... in the political campaigns so far this year?" We assigned participants who answered "very much interested", "somewhat interested", "not much interested" to high, intermediate, weak interest groups, respectively.[2] The number of participants assigned to each strength group at each election can be found in Table 4.1.

Global attitude measure In both the pre- and post-election interview, participants rated how warm and favorable they felt toward the presidential candidate on a scale from 0 to 100, with 0 representing a very unfavorable attitude toward the presidential candidate and with 100 representing a very favorable attitude toward the presidential candidate.

Voting decision. In the post-election interview, participants were asked which candidate they voted for. Depending on which presidential candidate the analysis focused, we scored the response as 1 when the par-

[1] As a check on the robustness of results, we also ran the analyses with answer option "moderately well" assigned to the second category and without dichotomization (by using polychoric correlations). The results of these alternative analyses mirrored the results reported in this chapter.

[2] In the ANES of 2008, the item was only administered to a subsample (N = 1178), while another subsample (N = 1144) was administered the item: "How interested are you in information about what's going on in government and politics?" The answer options were "Extremely interested", "Very interested", "Moderately interested", "Slightly interested", and "Not interested at all". Here, we assigned participants who answered either "extremely interested" or "very interested", "moderately interested", "slightly interested" or "not interested at all" to high, intermediate, weak attitude strength groups, respectively. The frequencies of the number of participants assigned to each group were similar for both interest items.

ticipants stated that they voted for the given candidate and we scored the response as 0 when the participants did not vote for the given candidate (cf., B. K. Payne et al., 2010).

4.6.3 Data Analysis

To estimate attitude network structures, we applied the eLasso-procedure (van Borkulo et al., 2014) to the responses on the non-behavioral evaluative reactions.[3] In the eLasso-procedure, each variable is regressed on all other variables, while the regression function is subjected to regularization to control the size of the statistical problem (see Friedman et al., 2008; Tibshirani, 1996). For each node, the set of edges that displays the best fit to the data is selected based on the fit of the regression functions according to the Extended Bayesian Information Criterion (Chen & Chen, 2008). Parameters are then based on the regression parameters in the selected neighborhood functions (van Borkulo et al., 2014). For each interest group at each election, we estimated networks for the non-behavioral evaluative reactions toward each candidate by using this procedure, which resulted in a total of 54 estimated networks. For further information on estimating and analyzing attitude networks, see Dalege, Borsboom, van Harreveld, and van der Maas (2017).

We used the Average Shortest Path Length (ASPL; West, 1996) as a measure of network connectivity. For every two given nodes in the network, the ASPL computes the length of the shortest path that connects the nodes and then averages these estimates. Dijkstra's algorithm was used to calculate path lengths (Brandes, 2001; Dijkstra, 1959; Newman, 2001). Dijkstra's algorithm minimizes the inverse of the distance between two node pairs using the weight of the edges (the absolute values of the regression parameters in the networks reported in this chapter). A low ASPL indicates high connectivity, while a high ASPL indicates low connectivity.

To investigate whether political interest predicts network connectivity, we fitted a general linear model with the interest groups as factor, the ASPL as dependent variable, and the number of nodes in the network as covariate to control for network size. For pairwise comparisons, we used Tukey's test that corrects p-values for multiple testing.

As a measure of the attitude's impact on behavior, we calculated the biserial correlation between the global attitude measure assessed before

[3]We did not include voting behavior in the estimation of the attitude networks in this analysis because (a) we wanted to use predictability of voting behavior as an attitude strength index and (b) voting behavior was not assessed at the same time as the other evaluative reactions.

the election and the voting decision. We estimated the attitude's impact on behavior for each interest group at each election and for each candidate. To test whether network connectivity predicts the attitude's impact on behavior we calculated the partial correlation between network connectivity and the attitude's impact on behavior on the interest group level with network size partialled out.

As a measure of the attitude's stability, we calculated the Pearson correlation between the global attitude measure assessed before and after the election. We estimated the attitude's stability for each interest group at each election and for each candidate. To test whether network connectivity predicts the attitude's stability we calculated the partial correlation between network connectivity and the attitude's stability on the interest group level with network size partialled out.

The vast majority of missing values in the pre-election interview were caused by participants answering "don't know" on items tapping evaluative reactions. We omitted variables that had more than 10% missing values to decrease the number of participants who had to be excluded, as we applied casewise deletion to cases with missing values.[4] The numbers of these excluded participants can be found in Table 4.1 and Table 4.2 shows which variables were omitted. Participants, who did not answer the post-election interview or who did not vote in the presidential election had to be excluded for the analyses on the attitude's impact on behavior and stability (we excluded participants only for the relevant analyses).

4.7 Results

Political interest strongly predicted connectivity of attitude networks (see Figure 4.2). Note that some of the low interest networks (Mondale 1984, Bush 1988, Kerry 2004) were not fully connected (i.e. not all nodes were directly or indirectly connected), which leads to infinitely large shortest path lengths between disconnected nodes. To be able to still enter such networks into the analysis, we set infinitely large shortest path lengths to the highest shortest path length in the same network that was a real number. Note that this technique results in overestimation rather than underestimation of a network's connectivity, which implies that this technique is conservative with respect to the focal hypothesis (i.e., leads to higher

[4]We also ran the analyses with inclusion of all variables, imputed missing values randomly with either 0 or 1 or scored "don't know" as a middle point of the scale, controlled for sample size and variance, and estimated networks using polychoric correlations. The results of these analyses mirrored the results reported in this chapter.

connectivity estimates in low interest groups).

The fitted linear model (including the number of nodes as a covariate) showed a significant effect of the interest groups on the ASPL of the networks, F (2, 50) = 17.59, p < .001, η_p^2 = .41. All groups differed significantly from each other in the expected direction and these differences were marked by high effect sizes. The mean ASPL of the intermediate interest groups (M = 2.07) was lower than the mean ASPL of the low interest groups (M = 2.44), t (50) = 3.43, p = .004, 95% CI [0.16; 0.58], d = 1.62, indicating that the networks of the intermediate interest groups had a higher connectivity than the networks of the low interest groups. The mean ASPL of the high interest groups (M = 1.80) was lower than the mean ASPL of the intermediate interest groups, t (50) = 2.48, p = .044, 95% CI [0.06; 0.48], d = 1.17, indicating that the networks of the high interest groups had a higher connectivity than the networks of the intermediate interest groups. The mean ASPL of the high interest groups was lower than the mean ASPL of the low interest groups, t (50) = 5.91, p < .001, 95% CI [0.43; 0.85], d = 2.78, indicating that the networks of the high interest groups had a higher connectivity than the networks of the low interest groups.

The finding that political interest predicts connectivity of attitude networks allowed us to test the central hypothesis of this chapter – that network connectivity predicts attitude strength. This hypothesis was strongly supported. Network connectivity (after controlling for the size of the network) was highly related to both the attitude's impact on behavior, r = -.71, p < .001, see Figure 4.3a, and to the attitude's stability, r = -66, p < .001, see Figure 4.3b. This indicates that network connectivity is related to both the durability and impact of attitudes.

4.8 Discussion

We applied network analysis to the archival data of the ANES 1980-2012. By focusing on attitudes toward presidential candidates, we investigated the central postulate of the Causal Attitude Network (CAN) model that strong attitudes correspond to highly connected attitude networks. We first showed that political interest strongly predicts network connectivity of attitudes. Because interest in attitude objects is related to but not a defining feature of attitude strength, we then tested whether connectivity of attitude networks predicts two central features of attitude strength – the attitude's impact on behavior and stability. Network connectivity strongly predicted both impact on behavior and stability, supporting the

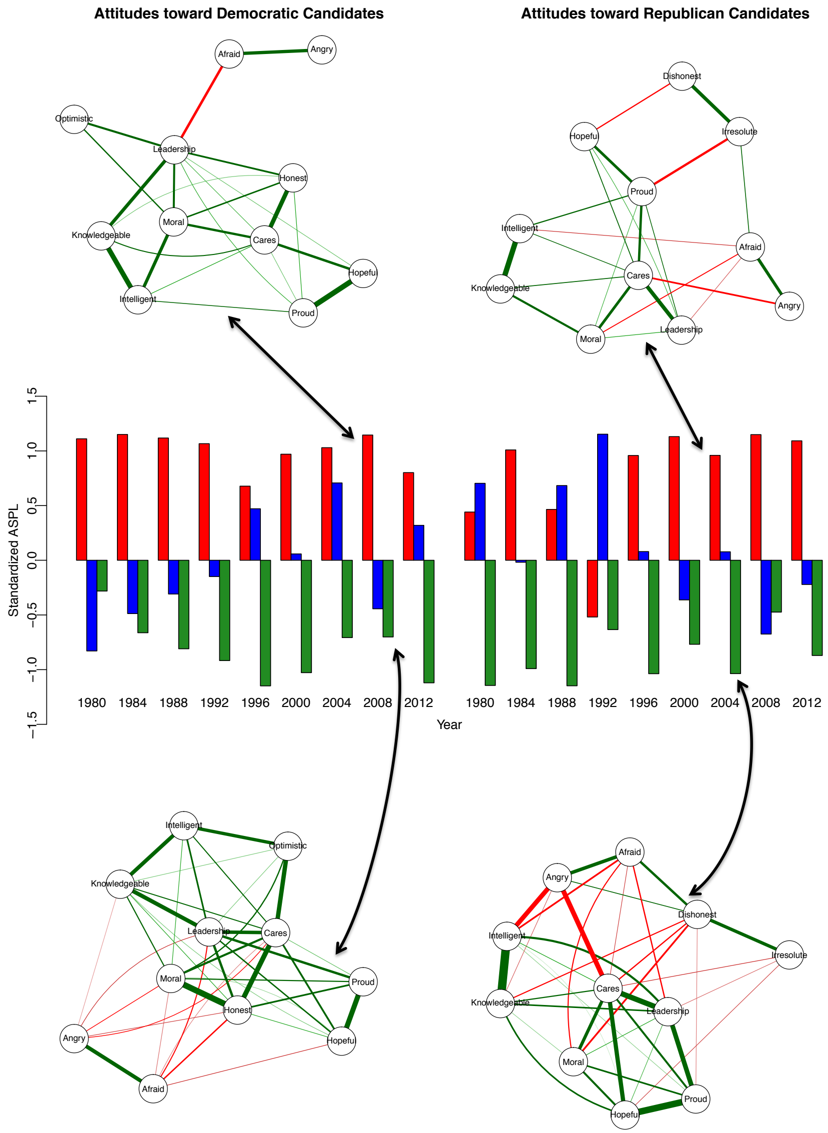


Figure 4.2: The barplot shows the standardized ASPL (number of standard deviations above or below mean for each set of interest groups) for each candidate at each year and for each interest group. Red (blue) [green] bars represent low (intermediate) [high] interest groups. Two representative networks of the low (high) interest groups are shown above (below) the barplot. The left (right) networks represent attitude networks toward Barack Obama (George W. Bush) in 2008 (2004). Nodes represent evaluative reactions (see Table 4.2 for the complete wording of the items), green lines represent positive connections, red lines represent negative connections, and thickness of an edge represents the strength of the connection. Closely connected nodes are placed near each other (Fruchterman & Reingold, 1991). All networks shown in this chapter were constructed using the R-package qgraph (Epskamp et al., 2012).

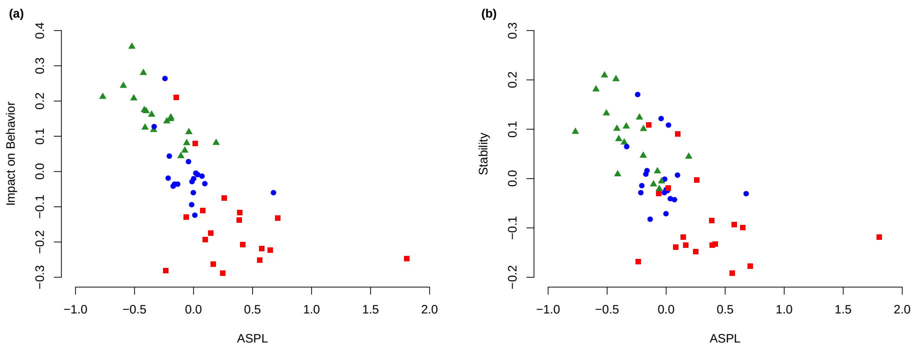


Figure 4.3: Relation between network connectivity (ASPL) and attitude strength (impact on behavior and stability). (a) Scatterplot of the relation between ASPL and attitude's impact on behavior (controlled for network size). (b) Scatterplot of the relation between ASPL and attitude stability (controlled for network size). Red squares (blue circles) [green traingles] represent low (intermediate) [high] interest groups.

central postulate of the CAN model that highly connected attitude networks correspond to strong attitudes.

4.8.1 Limitations

While using archival data from the ANES for our analyses has several advantages (e.g., large and representative samples), it also has some drawbacks. The measure of interest had a conceptually different focus (i.e., political campaigns) than the attitude measures (i.e., presidential candidates). However, a measure of interest more closely related to the measures of attitude is more likely to increase the relation between interest and network connectivity than to decrease it. We would also argue that during the election year, political campaigns are focused on the presidential candidates – making the measure of interest more closely related to the attitude measures.

Another limitation of the current study is that it only included measures of two of the four central features of attitude strength. Feature research therefore should investigate the relation between network connectivity and (1) resistance to persuasion and (2) impact on information processing. Additionally, the current study only measured one specific set of attitudes – attitudes toward presidential candidates. Future research might investigate the generality of the relation between network connectivity and attitude strength by focusing on a more diverse set of attitudes.

Given that the estimated networks were based on groups of individuals, it is not straightforward to generalize our findings to the level of the individual (e.g., Borsboom & Cramer, 2013). However, the group-based networks reported in this chapter are likely to be representative of individually based networks if the groups are relatively homogenous (e.g., individuals belonging to the same group have similar network structures; van Borkulo et al., 2015). It is likely that groups were homogenous because the assignment to political interest groups made the groups more homogenous and it is also likely that connections between different evaluative reactions differ more in quantity than in quality (e.g., the connection between judging a presidential candidate to be caring and judging a presidential candidate to be moral probably is positive for most individuals). Nonetheless, future research should investigate whether our findings also replicate at the individual level using appropriate techniques. The challenge here is that attitude networks currently can only be estimated based on several data points (either several data points based on several persons or based on several measurements per person). This makes it especially difficult for research at the individual level, because a highly connected network might limit the variation at the individual level, so that strong connections might not be recovered because of too low variation.

4.8.2 Implications and Directions for Future Research

Our results suggest that at least some strong attitudes are based on highly connected networks. This finding has fundamental implications for theorizing on attitude strength and attitude-behavior consistency. Linking attitude strength to network connectivity provides a novel and promising way to derive predictions regarding the dynamics of strong attitudes (that are based on highly connected networks). Such attitudes will generally be highly stable, but can also show instances of high instability under specific circumstances. This would be the case when the attitude network is highly connected but at the same time has to integrate a large amount of conflicting information.

Network connectivity also provides hypotheses regarding *how* attitudes change (Dalege et al., 2016). Change in weakly connected attitude networks takes place on a continuum ranging from a configuration in which all evaluative reactions are negative, to a configuration in which all evaluative reactions are positive, with unaligned configurations being only slightly less stable than aligned configurations. Change in highly connected attitude networks, in contrast, occurs more in an all-or-none fashion, with aligned configurations being much more stable than unaligned configurations. These dynamical characteristics of weakly versus highly connected attitude networks link our work to the catastrophe model of attitudes (Latané & Nowak, 1994; van der Maas et al., 2003; Zeeman, 1976), which holds important attitudes act like categories (i.e., attitudes can be either positive or negative), while unimportant attitudes act like dimensions (i.e., attitudes represent a continuous dimension running from positive to negative). Linking the connectivity hypothesis to the catastrophe model of attitudes leads to the prediction that important (unimportant) attitudes act like categories (dimensions) because they correspond to highly (weakly) connected networks.

Focusing on the connectivity of attitude networks provides novel opportunities for both predicting and influencing (voting) behavior (see also Dalege, Borsboom, van Harreveld, Waldorp, & van der Maas, 2017). Regarding behavior prediction, identifying whether individuals' attitude networks are densely or sparsely connected can inform researchers whether it is likely or not that the measured evaluative reactions will predict subsequent behaviors. Pollsters might benefit from that because this provides a novel opportunity to estimate how likely it is that the results of a given poll will translate into election results. Regarding influencing behavior, connectivity of attitude networks can inform which strategy should be taken to influence behavior. If the attitude network is weakly connected, a focus on the behavior itself may be a means to influence behavior effectively. If the attitude network is highly connected, however, the most promising way to influence behavior is probably to first apply strategies to decrease the connectivity of the attitude network (e.g., by lowering the importance of the attitude). When the connectivity has decreased, it is probably easier to induce the desirable behavior. To enhance longevity of attitude change, it would be beneficial to heighten the connectivity of the attitude network after the desired behavior is induced.

An interesting avenue for future research is also to investigate how network connectivity relates to other attributes of attitude strength, such as ambivalence, certainty or accessibility. While it is beyond the scope of the current chapter to discuss the relation between network connectivity and all attributes related to strength, we discuss some potentially fruitful avenues for future research on such relations. First, based on the CAN model (Dalege et al., 2016), we would argue that it is likely that network connectivity amplifies the relation between potential ambivalence (i.e., number of conflicting evaluations) and felt ambivalence (i.e., discomfort resulting from ambivalence, e.g., Priester & Petty, 1996). Second, knowledge on the attitude object was shown to amplify the effects of attitude strength (e.g., high knowledge in combination with high attitude strength leads to very pronounced stability and impact of attitudes). Similarly, network size (which represents a straightforward conceptualization of knowledge on the attitude object) amplifies the effects of network connectivity (Cramer et al., 2016). For a more comprehensive discussion of the relations between network connectivity and attitude strength, we refer the interested reader to Dalege et al. (2016).

4.8.3 Conclusion

In this chapter, we provided support for the connectivity hypothesis, which holds that highly connected attitude networks correspond to strong attitudes. The connectivity hypothesis provides several novel pathways for research into the different dynamics of strong and weak attitudes, such as instances of high instability of strong attitudes, indicating that network theory shows promise in providing a novel and fruitful framework of attitude strength.

Chapter 5

Network Structure Explains the Impact of Attitudes

Abstract

Attitudes can have a profound impact on socially relevant behaviors, such as voting. However, this effect is not uniform across situations or individuals, and it is at present difficult to predict whether attitudes will predict behavior in any given circumstance. Using a network model, we demonstrate that (a) more strongly connected attitude networks have a stronger impact on behavior, and (b) within any given attitude network, the most central attitude elements have the strongest impact. We test these hypotheses using data on voting and attitudes toward presidential candidates in the US presidential elections from 1980 to 2012. These analyses confirm that the predictive value of attitude networks depends almost entirely on their level of connectivity, with more central attitude elements having stronger impact. The impact of attitudes on voting behavior can thus be reliably determined before elections take place by using network analyses.

This chapter has been adapted from: Dalege, J., Borsboom, D., van Harreveld, F., Waldorp, L. J. & van der Maas, H. L. J. (2017). Network structure explains the impact of attitudes on voting decisions. *Scientific Reports*, *7*, 4909.

5.1 Introduction

Suppose you are one of the more than 130 million Americans who voted in the presidential election in 2016. Let us further assume that you were supportive of Hillary Clinton: You mostly held positive beliefs (e.g., you thought she was a good leader and a knowledgeable person) and you had positive feelings toward her (e.g., she made you feel hopeful and proud), representing a positive attitude toward Hillary Clinton (Dalege et al., 2016; Eagly & Chaiken, 1993; Fishbein & Ajzen, 1975; Rosenberg et al., 1960). However, you also held a few negative beliefs toward her (e.g., you thought that Hillary Clinton was not very honest). Did your overall positive attitude cause you to vote for Hillary Clinton? Here we show that the answer to this question depends on the network structure of your attitude: First, we show that the impact of attitudes (i.e., average of the attitude elements) on behavioral decisions depends on the connectivity of the attitude network (e.g., your network of your positive attitude toward Hillary Clinton was highly connected, so you probably voted for Hillary Clinton). Second, we show that central attitude elements have a stronger impact on behavioral decisions than peripheral attitude elements (e.g., your positive beliefs about Hillary Clinton were more central in your attitude network than your negative beliefs, so the chance that you voted for Hillary Clinton further increased). We thus provide insight into how structural properties of attitudes determine the extent to which attitudes have impact on behavior.

In network theory, dynamical systems are modeled as a set of nodes, representing autonomous entities, and edges, representing interactions between the nodes (Newman, 2010). The set of nodes and edges jointly defines a network structure. Modeling complex systems in this way has probably become the most promising data-analytic tool to tackle complexity in many fields (Barabási, 2011), such as physics (Barabási & Albert, 1999; Watts & Strogatz, 1998), biology (Barabási & Zoltán, 2004), and psychology (Cramer et al., 2010, 2012; van de Leemput et al., 2014; van Borkulo et al., 2015). Recently, network analysis has also been introduced to the research on attitudes in the form of the Causal Attitude Network (CAN) model (Dalege et al., 2016). In this model, attitudes are conceptualized as networks, in which nodes represent attitude elements that are connected by direct causal interactions (see Figure 5.1). The CAN model further assumes that the Ising model (Ising, 1925), which originated from statistical physics, represents an idealized model of attitude dynamics.

In the Ising model, the probability of configurations (i.e., the states of all nodes in the network), which represents the overall state of the attitude

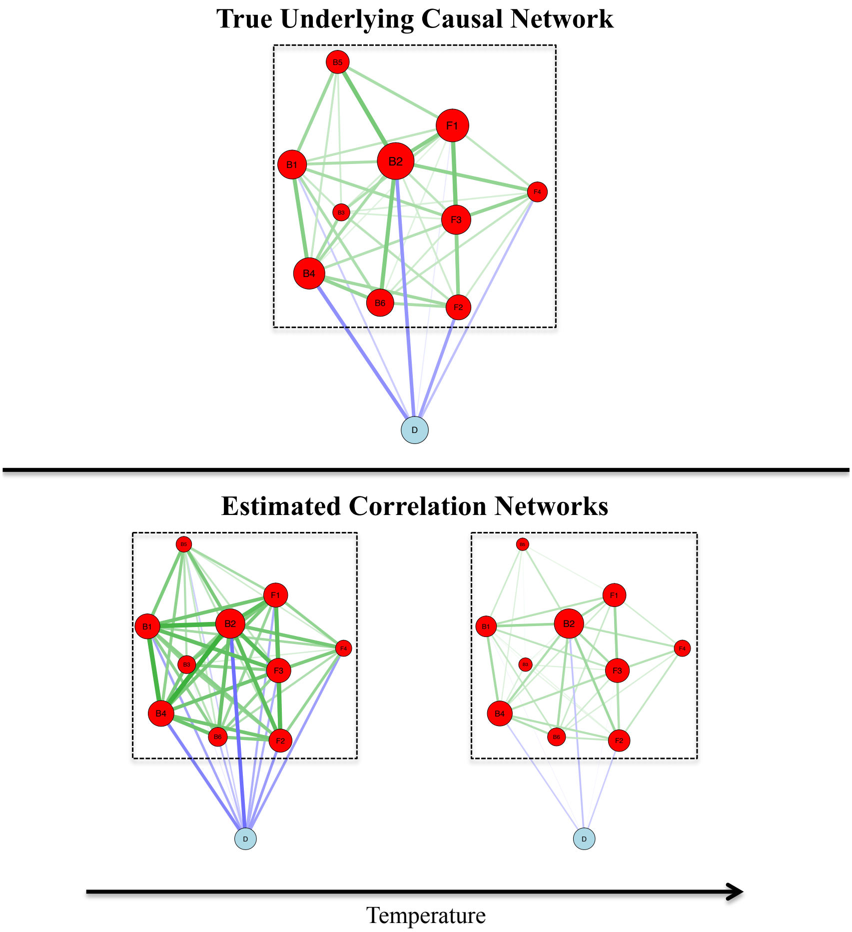


Figure 5.1: Illustrations of the Causal Attitude Network model and the hypotheses of the current study. Networks represent a hypothetical attitude network toward a presidential candidate consisting of six beliefs (e.g., judging the candidate as honest, intelligent, caring; represented by nodes B1 to B6), four feelings (e.g., feeling hope, anger toward the candidate; represented by nodes F1 to F4), and the voting decision (represented by the node D). Red nodes within the dashed square represent the part of the network on which connectivity and centrality estimates are calculated. Edges represent positive bidirectional causal influences (correlations) in the causal network (correlation networks), with thicker edges representing higher influence (correlations). Note that in this network, we assume that positive (negative) states of all nodes indicate a positive (negative) evaluation (e.g., positive state of judging a candidate as honest [dishonest] would be to [not] endorse this judgment). Size of the red nodes corresponds to their closeness centrality (see Methods for details on the network descriptives). In the CAN model, temperature represents a formalized conceptualization of consistency pressure on attitude networks. The correlation networks illustrate that lower (higher) temperature implies higher (lower) correlations between the attitude elements.

network, depends on the amount of *energy* of a given configuration. The energy of a given configuration can be calculated using the Hamiltonian function:

$$H(x) = -\sum_{i} \tau_i x_i - \sum_{<i,j>} \omega_{ij} x_i x_j. \tag{5.1}$$

Here, k distinct attitude elements $1,....,i,j,....k$ are represented as nodes that engage in pairwise interactions; the variables x_i and x_j represent the states of nodes i and j respectively. The model is designed to represent the probability of these states as a function of a number of parameters that encode the network structure. The parameter τ_i is the threshold of node i, which determines the disposition of that node to be in a positive state (1; endorsing an attitude element) or negative state (-1, not endorsing an attitude element) regardless of the state of the other nodes in the network (statistically, this parameter functions as an intercept). The parameter ω_{ij} represents the edge weight (i.e., the strength of interaction) between nodes i and j. As can be seen in this equation, the Hamiltonian energy decreases if nodes are in a state that is congruent with their threshold and when two nodes having positive (negative) edge weights assume the same (different) state. Assuming that attitude elements of the same (different) valence are generally positively (negatively) connected, attitude networks thus strive for a consistent representation of the attitude. The probability of a given configuration can be calculated using the Gibbs distribution (Murphy, 2012):

$$\Pr(X = x) = \frac{exp(-\beta H(x))}{Z}, \tag{5.2}$$

in which β represents the inverse temperature of the system, which can be seen as consistency pressures on attitude networks: reducing (increasing) the temperature of the system results in stronger (weaker) influence of the thresholds and weights, thereby scaling the entropy of the Ising network model (Epskamp, Maris, et al., 2018; Wainwright & Jordan, 2008). An Ising model with low (high) temperature results in a highly (weakly) connected correlation network (see Figure 5.1). The denominator Z represents the sum of the energies of all possible configurations, which acts as a normalising factor to ensure that the sum of the probabilities adds up to 1.

Conceptualizing attitudes as Ising models allows for the derivation of several hypotheses and a crucial test of this conceptualization is whether it can advance the understanding of the relation between attitudes and behavioral decisions. In the present chapter we apply the CAN model and are the first to (a) formalize and (b) test hypotheses based on the CAN model regarding the impact of attitudes on behavior.

The impact of attitudes on behavior has been one of the central research themes in social psychology in recent decades (Ajzen, 1991; Glasman & Albarracín, 2006; Kruglanski et al., 2015). The bulk of the research on the relation between attitudes and behavior has been done under the umbrella definition of attitude strength, which holds that one central feature of strong attitudes is that they have a strong impact on behavior (Krosnick & Petty, 1995). Several lines of research have identified factors related to attitude strength. Among the most widely researched of these are attitude accessibility, attitude importance, and attitudinal ambivalence. Studies have shown that accessible attitudes (i.e., attitudes that can be easily retrieved from memory) have more impact on behavior (Fazio & Williams, 1986; Glasman & Albarracín, 2006). Similarly, higher levels of (subjective) attitude importance (i.e., attitudes, to which a person attaches subjective importance), are related to increased accessibility of attitudes (Krosnick, 1989) and to higher levels of consistency between attitudes and behavior (Krosnick, 1988; Visser et al., 2003). Ambivalent attitudes (i.e., attitudes that are based on both negative and positive associations) are less predictive of behavior than univalent attitudes (Armitage & Conner, 2000; van Harreveld et al., 2015). While these and other attitude strength attributes, such as certainty and extremity, are generally interrelated (Krosnick et al., 1993; Visser et al., 2006), a framework that unifies these different attributes has long been absent in the literature. Recently, however, based on the development of the CAN model, attitude strength was formally conceptualized as network connectivity (Dalege et al., 2016). The CAN model might thus provide the basis for a comprehensive and formalized framework of the relationship between attitudes and behavior. Our current aim is to develop and test such a framework. To do so, we first formally derive hypotheses regarding the impact of attitudes on behavior from the CAN model. Second, we test these hypotheses in the context of voting decisions in the US American presidential elections.

From the CAN model the hypothesis follows that highly connected attitude networks (i.e., attitude networks that are based on Ising models with low temperature) have a strong impact on behavior. As can be seen in Figure 5.1, low temperature results in strong connections both between non-behavioral attitude elements (i.e., beliefs and feelings) *and* between non-behavioral attitude elements and behaviors (e.g., behavioral decisions, Dalege et al., 2016). Attitude elements in highly connected networks are thus expected to have a strong impact on behavioral decisions. This leads to the hypothesis that the overall impact of attitudes depends on the connectivity of the attitude network. While the connectivity of attitude networks provides a novel formalization of attitude strength, earlier approaches to understanding the structure of attitudes fit very well within this framework. For example, studies have shown that important attitudes are more coherent than unimportant attitudes (Judd et al., 1981; Judd & Krosnick, 1989) and that strong attitudes have a more consistent structure between feelings and beliefs than weak attitudes (Chaiken et al., 1995). Also, Phillip E. Converse's (1970) distinction between attitudes and nonattitudes based on stability of responses relates to our connectivity framework (Dalege et al., 2016)

In addition to predicting the overall impact of an attitude from the connectivity of the attitude network, the CAN model predicts that the specific impact of attitude elements depends on their centrality (as defined by their closeness). Closeness refers to how strongly a given node is connected both directly and indirectly to all other nodes in the network (Freeman, 1978; Opsahl et al., 2010). In contrast to connectivity, which represents a measure of the whole network, centrality is a measure that applies to individual nodes within the network. Attitude elements high in closeness are good proxies of the overall state of the attitude network, as they hold more information about the rest of the network than peripheral attitude elements, rendering closeness the optimal measure of centrality for our current purposes. We therefore expect central attitude elements to have a stronger impact (directly or indirectly) on a behavioral decision no matter which attitude elements are direct causes of this decision. This can also be seen in Figure 5.1, as there is a strong relation between a given node's centrality and it's correlation with the behavioral decision. It is important to note here that centrality of attitude elements does not refer to the classical definition of attitude centrality, but to the network analytical meaning of centrality. Specific impact of attitude elements has received somewhat less attention in the attitude literature than the global impact of attitudes, with studies either focusing on the primacy of feelings or beliefs in determining behavior (Galdi, Arcuri, & Gawronski, 2008; Lavine, Thomsen, Zanna, & Borgida, 1998; Millar & Millar, 1996) or on the subjective importance of attitude elements (van Harreveld et al., 2000; van der Pligt et al., 2000) and these different lines of research have been carried out much in isolation from each other and from the attitude strength research paradigm (for an exception see van Harreveld et al., 2000). It is our view that an advantage of the approach we take in this chapter is that our framework holds promise in unifying these different approaches to understanding the relation between attitudes and behavior.

In this chapter, we first show that the hypotheses put forward here above directly follow from conceptualizing attitudes as networks with a simulation study. We then test these hypotheses using data on attitudes toward candidates and voting in the American presidential elections from 1980–2012. In doing so, we test whether the CAN model provides a comprehensive framework on whether attitudes and which attitude elements drive behavioral decisions. Voting decisions are a perfect test of this postulate, because political attitudes often but not always drive voting decisions (Fazio & Williams, 1986; Galdi et al., 2008; Lavine et al., 1998; Kraus, 1986; Markus, 1982).

5.2 Results

5.2.1 Simulation Study

To show that the hypotheses presented above directly follow from conceptualizing attitudes as networks, we simulated networks using three popular algorithms to generate networks: preferential attachment (Barabási & Albert, 1999; Albert & Barabási, 2002), small-world network model (Watts & Strogatz, 1998), and random Erdos-Rényi networks (Erdos & Rényi, 1959), see also Supplementary Note 1 in Appendix D for analytical solutions. The networks consisted of 11 nodes (which corresponds to the number of nodes in the empirically estimated networks described below), with ten randomly chosen nodes representing attitude elements and one randomly chosen node representing the behavioral decision. Note that in such small networks, network properties other than density and magnitude of edge weights do not play a fundamental role in determining outcomes of the network.

The simulation of networks followed four steps: First, we created a 'base' network using one of the three algorithms. Second, we added edge weights to the base network, either drawn from a normal distribution, a Pareto power law distribution, or a uniform distribution. Third, to simulate responses of individuals holding attitudes with the network structure of the base network, we used the Ising network model (Ising, 1925). We created 20 different variations of the weighted base network in which the temperature of the Ising model was varied. Fourth, we simulated 1000 individuals based on the variations of the base network. As can be seen in Figure 5.1, increasing (decreasing) the temperature results in decreasing (increasing) edge weights in the correlation networks.

We repeated this procedure 100 times for each combination of network generating algorithms and edge weights distributions. To investigate whether simulated attitude elements in highly connected networks (i.e., networks, for which the temperature parameter was low) collectively have a strong impact on the simulated decision, we estimated the global connectivity, defined by the Average Shortest Path Length (ASPL, West, 1996) of the simulated attitude elements. We correlated the global connectivity with the average impact (which we operationalize as the biserial correlation between the sum score of the simulated attitude elements and the simulated decision) for each set of 20 networks. This resulted in strong negative correlations collapsed over all combinations of network-generating algorithms and edge weights distributions (Pearson correlations: mean r=-0.91, s.d. r=0.06) and we found strong negative correlations for all of these combinations (see Table D.1 in Appendix D). To investigate whether central nodes (based on closeness) have a strong impact on a decision, we estimated the centrality of the simulated attitude elements and correlated the centrality estimates with the impact of the simulated attitude elements (which we operationalize as the tetrachoric correlation between a given simulated attitude element and the simulated decision). To exclude the possibility that results are driven by differences in average centrality and impact, we standardized both centrality and impact for each network. This resulted in strong positive correlations in the different sets of attitude networks collapsed over all combinations of network-generating algorithms and edge weights distributions (Pearson correlations: mean r=0.59, s.d. r=0.29) and we found strong positive correlations for all of these combinations (see Table D.1 in Appendix D).

5.2.2 Test of Connectivity Hypothesis

These simulations show clearly that the CAN model predicts a strong relation between network connectivity (node centrality) and the predictive utility of attitudes (attitude elements) in forecasting behavior. This confirms that these intuitively derived hypotheses are indeed formal predictions that must follow if the CAN model is a valid model of attitudes. To provide an empirical test of the hypotheses put forward here, we analyzed data from the American National Election Studies (ANES) on the US presidential elections from 1980–2012 (total n=16,988). In each ANES between ten and 24 attitude elements were assessed and we selected ten attitude elements for each election that were most similar to each other, see Table 5.1. On these ten attitude elements, we estimated attitude networks for each of the two (three) main candidates for the elections in 1984–1992 and in 2000–2012 (in 1980 and 1996). This gave us 20 attitude networks in total. Nodes in these networks represent attitude elements toward the given

Table 5.1: Included Attitude Elements.

Attitude element	Included in data set	Substituted by
"is honest"*	1988–1996, 2008–2012	"is dishonest"* (1980, 2000–2004), "is decent"* (1984)
"is intelligent"*	1984–1992, 1996 (Clinton), 2000–2012	"is weak"* (1980), "gets things done"* (1996 Dole)
"is knowledgeable"*	1980–2012	NA
"is moral"*	1980–2012	NA
"really cares about people like you"*	1984–2012	"is inspiring"* (1980)
"would provide strong leadership"*	1980–2012	NA
"angry"**	1980–2012	NA
"afraid"**	1980–2012	NA
"hopeful"**	1980–2012	NA
"proud"**	1980–2012	NA

Note. *Denotes items tapping beliefs. Participants rated to which extent they agreed that these statements described the candidates. **Denotes items tapping feelings. Participants rated whether the candidates ever made them feel these feelings.

presidential candidate that were rated by the participants. Edges between the nodes represent zero-order polychoric correlations between the attitude elements. Note that because our networks are based on zero-order correlations, these networks only vary in magnitudes of edge weights and not in density, because correlation networks are always fully connected.

First, we tested whether highly connected attitude networks have strong average impact. As in the simulation study, connectivity was based on the ASPL and average impact was operationalized as the biserial correlation between the sum score of attitude elements and the voting decision. As can be seen in Figure 5.2, we found a high negative correlation between connectivity and average impact (Pearson correlation: r=-0.95, P<0.001), supporting our hypothesis.

5.2.3 Test of Centrality Hypothesis

Second, we tested whether central attitude elements have a strong impact. This impact was operationalized as the polychoric correlation between a given attitude element and the voting decision. We again standardized the centrality and impact estimates to exclude the possibility that results are driven by differences in mean centrality and mean impact. As can be seen in Figure 5.3, we found a high positive correlation between standardized centrality and standardized impact (Pearson correlation: r=0.70, P<0.001), supporting our hypothesis.

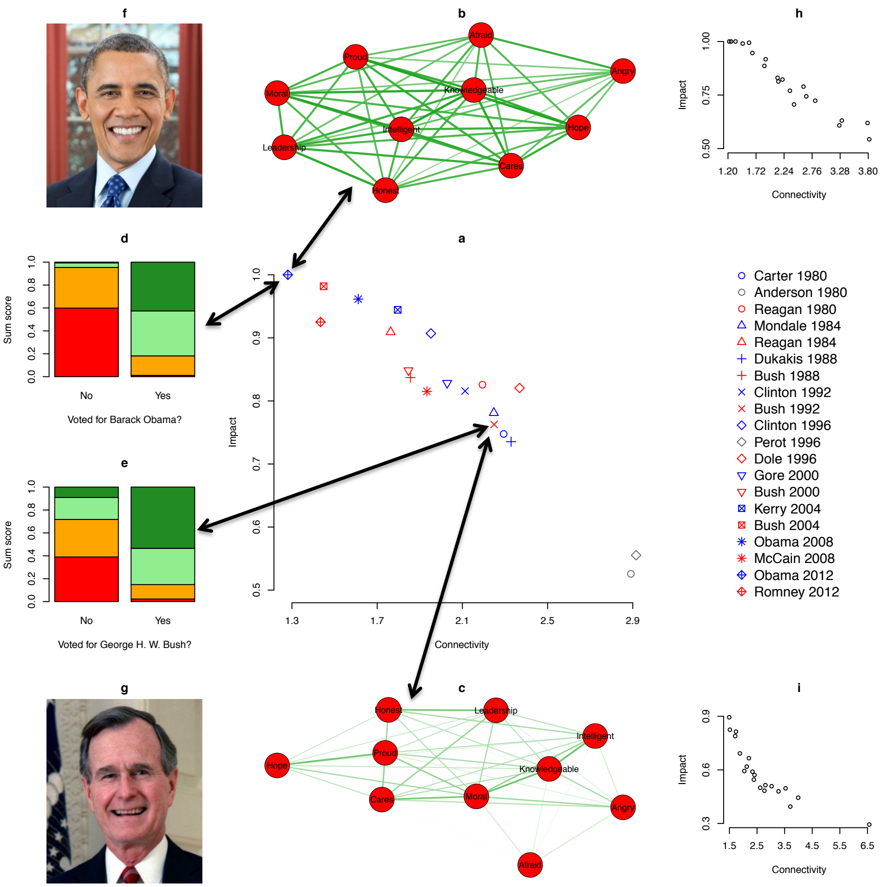


Figure 5.2: Highly connected attitude networks have a stronger impact on voting decisions than weakly connected attitude networks. (**a**) Relation between connectivity and average impact of attitude elements. (**b**)-(**e**) Two illustrations of the analytic strategy to assess connectivity and average impact. (b) [(c)] Attitude network toward Barack Obama [George H. W. Bush] in 2012 [1992]. Nodes represent attitude elements and edges represent correlations between attitude elements (the higher the correlation, the thicker the edge; correlations lower than .3 are not displayed). Closely connected attitude elements are placed near to each other (Fruchterman & Reingold, 1991). (**d**) [(**e**)] Relation between the sum score of attitude elements toward Barack Obama [George H. W. Bush] and voting for Barack Obama [George H. W. Bush]. Colors of the bars represent the percentage of individuals who's sum scores fall into a given percentile (the more green, the higher the sum score; the more red, the lower the sum score). (**f**) Photo of Barack Obama by Pete Souza. Photo is under the CC0 / Public Domain Licence. Source: https://www.goodfreephotos.com/people/barack-obama-portrait-photo.jpg.php. (**g**) Photo of George H. W. Bush by unknown photographer. Photo is under the CC0 / Public Doman Licence. Source: https://www.goodfreephotos.com/people/george-bush-portrait-photo.jpg.php. (**h**) [(**i**)] Relation between connectivity and impact for the simulated set of networks that was closest to the mean correlation plus [minus] one standard deviation.

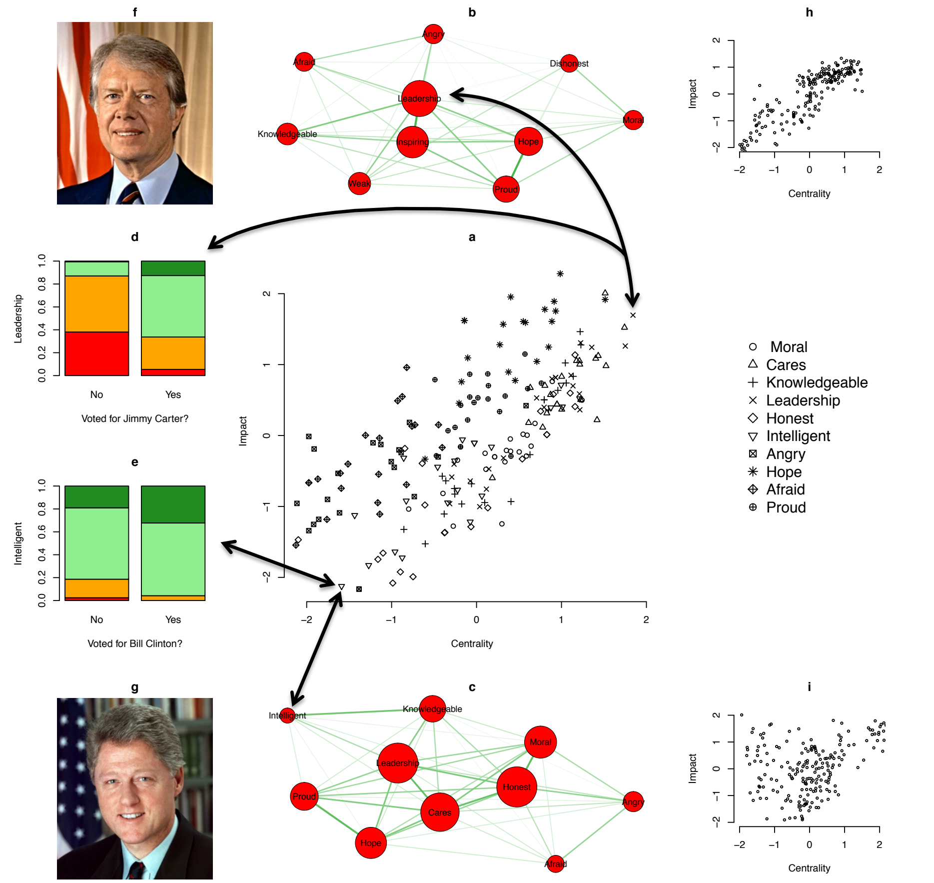


Figure 5.3: Central attitude elements have stronger impact on voting decisions than peripheral attitude elements. (**a**) Relation between centrality and impact of attitude elements. (**b**)-(**e**) Two illustrations of the analytic strategy to assess centrality and impact. (**b**) [(**c**)] Attitude network toward Jimmy Carter [Bill Clinton] in 1980 [1992]. The networks have the same characteristics as the networks shown in Figure 5.2, except that the size of the nodes corresponds to the nodes' relative centrality (the bigger the node, the higher its centrality). (**d**) [(**e**)] Relation between endorsing the belief that Jimmy Carter [Bill Clinton] would provide strong leadership [is intelligent] and voting for Jimmy Carter [Bill Clinton]. Colors of the bars represent the percentage of individuals who agree or do not agree with the judgment (the more green, the higher the agreement; the more red, the lower the agreement). See Table 5.1 for more information on the attitude elements. (**f**) Photo of Jimmy Carter by unknown photographer. Photo is under the CC0 / Public Domain Licence. Source: https://www.goodfreephotos.com/people/jimmy-carter-portrait.jpg.php. (**g**) Photo of Bill Clinton by Bob McNeely. Photo is under the CC0 / Public Doman Licence. Source: https://www.goodfreephotos.com/people/bill-clinton-portrait-photo.jpg.php. (**h**) [(**i**)] Relation between centrality and impact for the simulated set of networks that was closest to the mean correlation plus [minus] one standard deviation.

5.2.4 Forecast Analysis

To illustrate the practical relevance of our findings, we investigated whether centrality of attitude elements can be used to forecast the impact on voting decisions *before* knowing the outcome of the election (e.g., whether our analyses can be used to forecast the impact of attitude elements on the next presidential election). For each election (e.g., election of 2012), we estimated the regression parameters between impact and centrality from all elections except the forecasted election (e.g., 1980-2008). We calculated the predicted impact in the forecasted election using the centrality indices of the forecasted election and the regression parameters. As can be seen in Figure 5.4, the predicted impact was very close to the actual impact (deviation median=0.06, deviation interquartile range=0.03-0.09) and outperformed both using the mean of all attitude elements (Deviation median=0.12, deviation interquartile range=0.06-0.18, Wilcoxon-matched pairs test: V=3346, P<0.001, CLES=69.5%) and using the means of the specific attitude elements (Deviation median=0.09, deviation interquartile range=0.04-0.17, Wilcoxon-matched pairs test: V=5057, P<0.001, CLES=65.2%). Using centrality thus creates the possibility to forecast the (almost) exact impact of an attitude element on the voting decision.

5.3 Discussion

Starting in the 1930s with Richard T. LaPiere's (1934) work, attitude-behavior consistency has been one of the central research themes in social psychology (Ajzen, 1991; Fazio & Williams, 1986; Fishbein & Ajzen, 1975; Glasman & Albarracín, 2006; Kraus, 1986; Wicker, 1969). While early work focused on the question *whether* attitudes drive or do not drive behavior (Wicker, 1969), more recent work has focused on *when* attitudes drive behavior (Glasman & Albarracín, 2006; Bagozzi & Baumgartner, 1986; Ajzen & Fishbein, 1977; Fazio & Williams, 1986; Fazio & Zanna, 1981). This chapter provides a formalized and parsimonious answer to this question: The impact of attitudes on behavior depends on the connectivity of the attitude network, with central attitude elements having the highest impact on behavior within a given attitude network.

The present research has shown that network structure of attitudes can inform election campaign strategies (and behavioral change programs in general) by predicting both the extent to which individuals base their decision on their attitude and the extent to which an attitude element influences the voting decision (and other behavior relevant to an attitude).

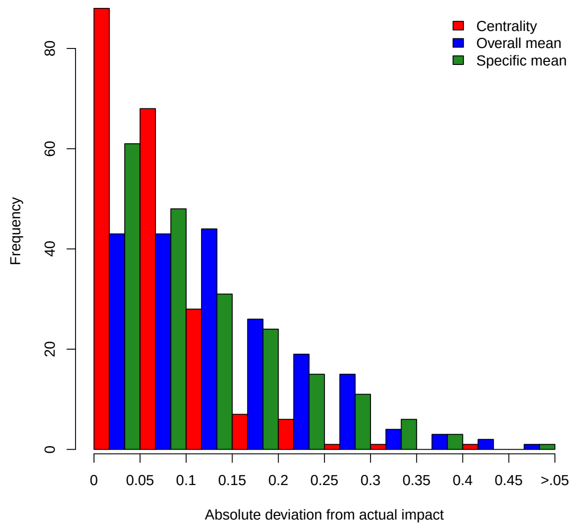


Figure 5.4: Accuracy of forecasts based on centrality, overall mean, and specific mean. The plot shows the results of forecasting the impact of each attitude element at each election.

Connectivity can help inform how effective candidate-centred campaigns would be. High connectivity indicates that voting decisions highly depend on candidate attitudes, while low connectivity indicates that other factors may play a more substantial role than candidate attitudes, such as party identification (L. M. Bartels, 2000; Miller, 1991), ideology (Jacoby, 2010; Palfrey & Poole, 1987), and public policy issues (Abramowitz, 1995; Aldrich, Sullivan, & Borgida, 1989; Carmines & Stimson, 1980; Palfrey & Poole, 1978; Nadeau & Lewis-Beck, 2001; Rabinowitz & MacDonald, 1989). Centrality can furthermore inform on the effectiveness of targeting specific attitude elements, as changing a central attitude element is probably more likely to affect the voting decision than changing a peripheral attitude element.

Future research might focus on how connectivity and centrality of attitude networks relate to other factors that influence voting decision. Among the most important factors influencing voting decisions are party identification (L. M. Bartels, 2000; Miller, 1991) and specific policy issues (Abramowitz, 1995; Aldrich et al., 1989; Carmines & Stimson, 1980; Palfrey & Poole, 1978; Nadeau & Lewis-Beck, 2001; Rabinowitz & MacDonald, 1989). First, party identification might influence the connectivity of attitude networks, because it is likely that individuals, who identify with a political party, have a stronger drive for consistency in their attitudes toward presidential candidates. Party identification makes it also more likely that a given individual adopts a positive attitude toward the candidate of their party and it might also directly influence the voting decision. This makes party identification a possible confound of our results and we therefore also ran our analyses including only individuals, who do not identify with a political party. The results of this analysis mirrored the results reported in this chapter (see Supplementary Note 3 & Figure D.3 in Appendix D). Second, policy issues might influence the centrality of attitude elements. If, for example, the current political climate is highly focused on foreign policies (e.g., the conflict in Syria), judging a candidate to be competent in respect to foreign policy making might take a central place in the attitude network. Generally, it is an important question for future research why some attitude elements are more central than others. Our analyses indicate that there are some attitude elements that are chronically central (see Figure 5.3), with some variation that might be due to the specifics of the political climate during the different elections.

Another promising venture for future research would be to investigate how attitude networks develop during an election campaign. To do so, one could apply several intermediate assessments during the election campaign (Erikson & Wlezien, 1999; Taleb, 2017; Wlezien, 2003). The use of such intermediate assessment was shown to improve the prediction of election outcomes (Taleb, 2017). How might attitude networks change during an election? Based on the CAN model, we expect that (a) the connectivity of attitude networks heighten during an election campaign and (b) attitude networks probably grow due to the addition of newly formed attitude elements (Dalege et al., 2016). Also, predictions regarding the success of an election campaign to change a given person's attitude can be derived from the CAN model. Individuals holding attitudes that are based on highly connected networks already at the beginning of an election campaign are likely to not change their attitudes. Election campaigns might thus benefit from focusing on individuals holding attitudes that are based on weakly connected networks (Dalege et al., 2018b).

In a broader sense, the CAN model advances our understanding of the relation between attitudes and behavioral decisions. Because the CAN model is a general model of attitudes, the results reported here likely generalize to other attitudes and behavioral decisions than those studied here as well. Using connectivity of attitude networks and centrality of attitude elements may for example provide more insight into issues such as which factors drive individuals to continue or stop smoking, buy a certain product, or behave aggressively toward a minority group. Furthermore, connectivity of attitude networks might unify the different approaches to explain variations in attitude-behavior consistency, as it is likely that network connectivity is the glue that holds these factors together (Dalege et al., 2016) and because our results indicate that network connectivity comprehensively explains variations in attitude-behavior consistency. Several predictions above and beyond the findings reported here can also be derived from the network structure of attitudes. For example, network structure predicts when and which persuasion attempts will be successful (Dalege et al., 2016). Network theory thus holds great promise for advancing our understanding of the dynamical and structural properties of attitudes and their relation to a plethora of consequential human behaviors.

5.4 Methods

5.4.1 Simulation of Networks

The simulation of networks followed four steps. First, an unweighted 'base' network consisting of 11 variables was created based on preferential attachment (Barabási & Albert, 1999; Albert & Barabási, 2002), the small-world network model (Watts & Strogatz, 1998), or the Erdos-Rényi random graph model (Erdos & Rényi, 1959) using the R package iGraph (Csárdi & Nepusz, 2006). The preferential attachment algorithm starts with one node and then adds one node in each time step. The probability to which nodes the new node connects depends on the degree of the old node:

$$\Pr(i) = \frac{k(i)^{\alpha} + 1}{\sum_i k(j)^{\alpha} + 1}, \tag{5.3}$$

where $k(i)$ is the degree of a given node. α was set to vary uniformly between 0.30 and 0.70. At each time step m edges were added to the network. m was set to vary uniformly between 4 and 6 (resulting in relatively dense networks, as was shown to be the case for attitude networks, Dalege et al., 2016). The small-world network model starts with a ring lattice with nodes being connected to n neighbours and then randomly rewires edges with a p probability. n was set to uniformly vary between 3 and 4 and p was set to uniformly vary between 0.05 and 0.10. In the Erdos-Rényi graph, nodes are randomly connected by a given number of edges. Number of edges was set to uniformly vary between 30 and 45.

Second, edge weights were added to the base network. To have psychometrically realistic edge weights, we drew edge weights from either a normal distribution with M=0.15 and SD=.0075, a Pareto power law distribution with α=3 and β=0.10, or a uniform distribution with range of 0.01–0.30.

Third, we created 20 variations of the weighted base network, in which the temperature of the Ising model was varied. The inverse temperature parameter β was drawn from a normal distribution with M=1 and SD=0.2 (with higher numbers representing low entropy). To ensure that all nodes have roughly the same variance, we drew thresholds of nodes from a normal distribution with M=0 and SD=0.25.

Fourth, using the R-package IsingSampler (Epskamp, Maris, et al., 2018), 1000 individuals for each of the variations of the base network were simulated based on the probability distribution implied by the Ising model. This procedure was repeated 900 times and each set of 20 variations of the different 900 base networks was analyzed separately.

5.4.2 Participants

The open-access data of the ANES involves large national random probability samples. Data were each collected in two interviews - one before and one after each presidential election from 1980 to 2012 - by the Center for Political Studies of the University of Michigan. In total, 21,365 participants participated in these nine studies (for *Ns* per study see Supplementary Table D.2), of which 16,667 participants stated that they voted for president. Non-voters were excluded from the analyses, because we assume that the decision whom to vote for is more likely to be part of the attitude network than the decision whether to vote or not. In Supplementary Note 2 in Appendix D, however, we show that similar findings are obtained when non-voters are included in the analysis.

5.4.3 Measures

In each of the studies between six and 16 items tapping beliefs and between four and eight items tapping feelings toward the presidential candidates were assessed in the pre-election interviews. Feelings were assessed on two-point scales and beliefs were assessed on four-point scales (in a subsample of the ANES of 2008 and in the ANES of 2012, beliefs were assessed on a five-point scale). To have comparable attitude networks between the different elections, we always used six items tapping beliefs and four items tapping feelings (see Table 5.1 for a list of included attitude elements). In the post-election interview, participants were asked which candidate they voted for. Depending on which presidential candidate the analysis focused, we scored the response as 1 when the participant stated that they voted for the given candidate and we scored the response as 0 when the participant did not vote for the given candidate.

5.4.4 Statistical analyses

We performed the same statistical analyses on the simulated and empirical data.

Network Estimation. Attitude networks were estimated using zero-order polychoric (tetrachoric) correlations between the (simulated) attitude elements as edge weights. We chose to use zero-order correlations as edge weights instead of estimating direct causal paths between the attitude elements because our simulations have shown that attitude networks based on zero-order correlations perform better than techniques that provide an estimate of the underlying causal network (van Borkulo et al., 2014).

Network descriptives. Both the ASPL and closeness are based on shortest path lengths (d) between nodes. To calculate shortest path lengths, we used Dijkstra's algorithm (Dijkstra, 1959), implemented in the R package qgraph (Epskamp et al., 2012):

$$d^w(i,j) = min(\frac{1}{w_{ih}} + \frac{1}{w_{hj}}). \tag{5.4}$$

ASPL is then the average of the shortest path lengths between each pair of nodes in the network. Closeness (c) was calculated using the algorithm for weighted networks developed by Opsahl et al. (2010), using the R package qgraph:

$$c(i) = \left[\sum_{j}^{N} d(i,j)\right]^{-1}. \tag{5.5}$$

Impact estimates. To estimate average impact of (simulated) attitude elements on (simulated) voting decisions, we calculated the biserial correlation between the sum score of (simulated) attitude elements and the (simulated) voting decision. We then calculated the Pearson correlation between connectivity and average impact for the 20 networks in the empirical study and for each set of 20 variations of the base networks in the simulation study. The clearly linear relation between connectivity and impact justified the use of Pearson correlation and significance testing. For the simulation study, we calculated the mean and standard deviation of the correlations obtained for each set of variations of the base network.

To estimate the impact of a given (simulated) attitude element on the (simulated) voting decision, we calculated the zero-order polychoric correlation between a given (simulated) attitude element and the (simulated) voting decision. We then calculated the Pearson correlation between standardized centrality and standardized impact of the attitude elements in the empirical study and for each set of 20 variations of the base networks in the simulation study. The clearly linear relation between centrality and impact justified the use of Pearson correlation and significance testing. For the simulation study, we calculated the mean and standard deviation of the correlations obtained for each set of variations of the base network.

Forecast analysis. For the forecast analysis, we first conducted nine regression analyses, in which impact was regressed on centrality. In each regression analysis, the forecasted election was omitted. From each of the regression equations, we first extracted the beta and intercept coefficients. Second, we multiplied the centrality indices of the forecasted election with the beta coefficient and added the intercept coefficient. Note that not in every election the same attitude elements were assessed. Of the ten used attitude elements, seven were assessed at each election. For the remaining three, we grouped the attitude elements together that were most similar to each other (see Table 5.1). Third, we compared the resulting estimates with the actual impact of the attitude elements and calculated the absolute deviance scores. We then compared the performance of the centrality prediction to the overall mean prediction and the specific mean prediction. For both these predictions, we again calculated predictions nine times, omitting one of the elections each time. For the overall mean prediction, we calculated the mean of all attitude elements and for the specific mean prediction, we calculated the mean of each specific attitude element. We tested whether the centrality prediction performed better than the overall mean prediction and the specific mean prediction using Wilcoxon signed-rank tests.

Missing values. Missing values were deleted casewise (Table D.2 in Appendix D shows the number of excluded participants per attitude network). Most missing values stemmed either from participants responding to an item that they did not know the answer or from non-participation during the post-election interview. Few missing values stemmed from interview errors.

Alternative analyses. We also ran several alternative analyses that confirmed the robustness of our results: We ran alternative analyses on non-voters (see Supplementary Note 2 & Figure D.2 in Appendix D), on independents (see Supplementary Note 3 & Figure D.3 in Appendix D), on missing values (see Supplementary Note 4 & Figure D.4 in Appendix D), on networks with different number of nodes (see Supplementary Note 5 & Figure D.5 in Appendix D), and on latent variable models (see Supplementary Note 6 & Table D.3 in Appendix D).

Part III

The Attitudinal Entropy (AE) Framework as a General Theory of Attitude

Chapter 6

The Attitudinal Entropy (AE) Framework

Abstract

This chapter introduces the Attitudinal Entropy (AE) framework, which builds on the Causal Attitude Network (CAN) model that conceptualizes attitudes as Ising networks. The AE framework rests on three propositions. First, attitude inconsistency and instability are two related indications of attitudinal entropy, a measure of randomness derived from thermodynamics. Second, energy of attitude configurations serves as a local processing strategy to reduce the global entropy of attitude networks. Third, directing attention to and thinking about attitude objects reduces attitudinal entropy. We first discuss several determinants of attitudinal entropy reduction and show that several findings in the attitude literature, such as the mere thought effect on attitude polarization and the effects of heuristic versus systematic processing of arguments, follow from the AE framework. Second, we discuss the AE framework's implications for ambivalence and cognitive dissonance.

This chapter has been adapted from: Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2018). The Attitudinal Entropy (AE) framework as a general theory of individual attitudes. *Psychological Inquiry, 29*, 175-193.

6.1 Introduction

A century of research on attitudes has produced an impressive amount of empirical findings and identified an abundance of concepts and processes related to attitudes. An important next step towards a thorough understanding of attitudes would be a theoretical framework able to explain these empirical findings from few first principles. The aim of this chapter is to develop such a framework. To do so, we make use of analogical modeling (Haig, 2005): We use statistical mechanics as a starting point for our framework, because of its advanced theory and because our earlier analysis has shown that a formalized measurement model of attitudes can be based on statistical mechanics principles (Dalege et al., 2016), and show that an analogous theoretical approach to attitude can explain a wide variety of empirical phenomena.

Statistical mechanics revolves around three fundamental properties of a system – entropy (a measure of the system's randomness), energy, and temperature[1]. To investigate whether statistical mechanics represents a fruitful starting point for a general theory of attitudes, we search for analogies of these fundamental properties and test whether the consequences of these analogies match empirical findings in the attitude literature. Based on this approach we derive the Attitudinal Entropy (AE) framework, which rests on three propositions. First, inconsistency and instability of an attitude represents attitudinal entropy and is therefore the natural state of an attitude. Second, the energy of the attitudinal representation serves as a local processing possibility to evaluate the global entropy of an attitude. Third, attention and thought directed at the attitude object have an analogous effect on the attitudinal representation as (inverse) temperature has on thermodynamic behavior – heightened attention and thought make attitudinal representations low in energy more likely and therefore reduce the entropy of the attitude.

The structure of this chapter is as follows. First, we discuss the main tenets of the AE framework. Second, we discuss determinants of reduction of attitudinal entropy and show that several findings in the attitude literature, such as individual vs. group effects of implicitly measured attitudes, the mere thought effect, and systematic vs. heuristic processing, follow from these determinants. Third, we discuss ambivalence (e.g., Priester & Petty, 1996) and cognitive dissonance (Festinger, 1957) from the perspective of the AE framework. Throughout the subsequent sections we

[1] Temperature, strictly speaking, might not be regarded as a fundamental property, because it can be derived from the relation between entropy and energy. However, for our current purposes it is beneficial to treat temperature as a fundamental property.

model several established phenomena in the attitude literature to show that the AE framework indeed holds promise in explaining several phenomena with few first principles. We also identify several predictions that can be derived from the AE framework in each discussion of a given phenomenon to illustrate the predictive power of the AE framework and to define an empirical agenda for future research. We close by discussing potential neural substrates of the AE framework's propositions, the AE framework's relation to other broad models of attitude, and several open questions that need to be addressed to further develop the AE framework.

6.2 The Attitudinal Entropy (AE) Framework

In this section, we discuss the meaning of attitudinal entropy and its implications for the dynamics of attitudes. We first discuss micro- and macrostates of attitudes and then turn to the meaning of attitudinal entropy. Based on these definitions, we derive the Attitudinal Entropy (AE) framework.

6.2.1 Attitudinal Micro- and Macrostates

The first question that needs to be addressed before we can define attitudinal entropy is what constitutes microstates and macrostates of an attitude. In statistical mechanics, a microstate refers to the microscopic configuration of a given system (e.g., the position of each oxygen molecule in the room you are sitting in) and a macrostate refers to the macroscopic behavior of a given system (e.g., whether all the oxygen molecules are centered in one corner or whether they are evenly dispersed throughout the room). In line with several theories on attitudinal structure (e.g., Dalege et al., 2016; Eagly & Chaiken, 1993; Fishbein & Ajzen, 1975; Rosenberg et al., 1960), we define the microstate of an attitude as *the configuration of the relevant beliefs, feelings, and behaviors toward an attitude object (i.e. attitude elements)*. As an example take the attitude toward snakes. The microstate of this attitude can be represented like this: attitude element 1 (e.g., snakes maintain ecological order) is positive, attitude element 2 (e.g., snakes are scary) is negative, attitude element 3 is negative (e.g., I run away when I see a snake), and so forth. The macrostate of an attitude is then defined as *the combination of all attitude elements (e.g., how many attitude elements are negative and how many are positive)*. Based on several theories on the integration of attitude elements into a global evaluation (e.g., N. H. Anderson, 1971; Cacioppo et al., 1989; Cunningham & Zelazo, 2007; Fazio, 1995; Zanna & Rempel, 1988), we assume that the global evaluation of an attitude object is strongly related to the macrostate of an attitude, in that it represents a context-depended weighted sum score of the attitude elements. Thus, we propose the following three definitions:

Definition 1: The configuration of the attitude elements constitutes the microstate of the attitude.

Definition 2: The number of positive versus negative attitude elements constitutes the macrostate of an attitude.

Definition 3: A situation-depended weighted sum score constitutes the global evaluation of an attitude object.

6.2.2 Attitudinal Entropy

Entropy is a concept originating from thermodynamics, where it was originally defined as energy that is lost when energy is transformed (e.g., from chemical to kinetic energy). Take as an example the situation when you walk up a steep hill. To do this, your body has to transform chemical energy in the form of calories to kinetic energy so that your legs move up the hill. However, during this transformation of energy, some energy is inevitably lost that is not put to work, constituting heat loss or entropy. This notion also lies at the heart of the second law of thermodynamics, which states that entropy of an isolated system *always* increases.

While the concept of entropy originated in classical thermodynamics, its application to statistical mechanics resulted in a broader use of entropy as a general measure of disorder or uncertainty in a system. The physicist Ludwig Boltzmann (1877) developed the statistical mechanics definition of entropy, which holds that a macrostate that can be realized by many microstates has higher entropy than a macrostate that can be realized by few microstates (see Figure 6.1a). As an example, take the distribution of oxygen molecules in the room you are sitting in right now. Luckily, the macrostate of the oxygen molecules being distributed evenly throughout the room can be realized by many more microstates (thus having higher entropy) than the macrostate of the oxygen molecules clustering at one position in the room.[2] As an intuitive example of why the likelihood of a macrostate depends on its Boltzmann entropy, imagine a simple slot machine with three fields that can either show a lemon, a peach, or a banana. The macrostate ”win” (i.e., all fields showing the same fruit) can then be realized by three microstates (e.g., three lemons). The macrostate ”lose” (i.e., the fields show at least two different fruits) on the other hand can be

[2]Otherwise you might get crushed by all oxygen molecules distributed at the position of the room you are in or you might suffocate because all oxygen molecules are at a different position than you.

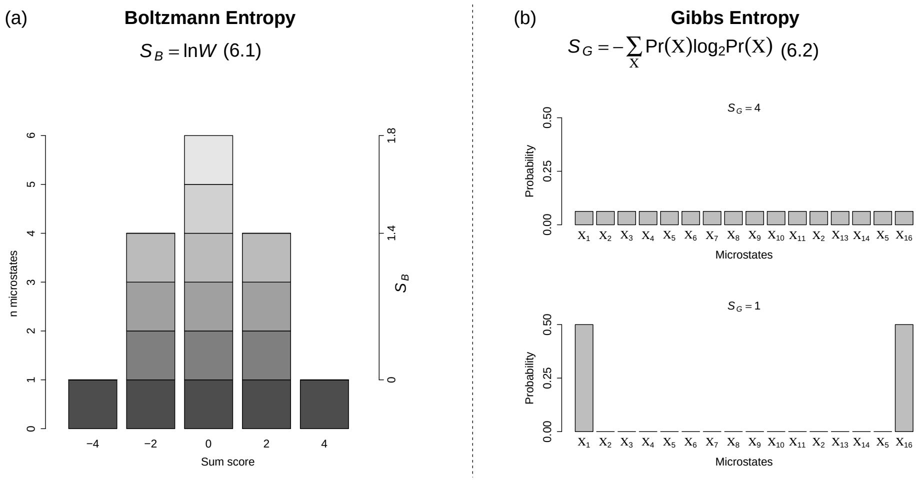


Figure 6.1: Illustrations of the Boltzmann and Gibbs entropies. In (1) W refers to the number of microstates that can realize the given macrostate. In (2) X refers to all possible states of a given system.

realized by 24 ($3^3 - 3$) microstates. While we already see with this simple example that high entropy states are more likely than low entropy states, the effect becomes more and more pronounced with the size of the system increasing (up to the point where the high entropy state is essentially the only possible state as is the case for the distribution of oxygen molecules in a room).

Applying the Boltzmann entropy to the domain of attitudes implies that inconsistent attitudes have higher entropy than consistent attitudes. To illustrate this, consider an attitude consisting of ten attitude elements. A perfectly univalent attitude can only be realized by two different microstates (i.e., all attitude elements being either positive or negative). So, the attitude of a snake enthusiast (i.e., judging snakes as entirely positive) can only be realized by one microstate. A perfectly ambivalent or neutral attitude, in contrast can be realized by 252 microstates. So, judging snakes as positive on some aspects and negative on others can be realized by a large number of microstates. This leads to the following first proposition of the AE framework:

Proposition I.1: Inconsistency of an attitude *is* the Boltzmann entropy of the attitude.

Important to note here is that the Boltzmann entropy concerns the entropy of a single given macrostate (e.g., five attitude elements are in a positive state and five attitude elements are in a negative state). The entropy of a system, on the other hand, is described by the Gibbs entropy

(Jaynes, 1965). Gibbs entropy depends on the likelihood of the different microstates of a system. As Figure 6.1b illustrates Gibbs entropy is at maximum when all microstates are equally likely – implying that the system's behavior is completely random – and it is at minimum when only a single configuration is possible – implying that the system's behavior is completely ordered. As an example of Gibbs entropy take the movement of water molecules. Under high temperature, water molecules move randomly (i.e., water is in a gas state) – this indicates high Gibbs entropy, because the configuration (i.e., positions) of the water molecules is consistently changing (i.e., all microstates are roughly equally likely). In contrast, under low temperature the water molecules cannot move (i.e., water is in a solid state) – this reflects low Gibbs entropy, because the configuration of the water molecules is stable (i.e., the current microstate is much more likely than all other microstates). Someone, who consistently changes her attitude toward snakes would therefore have a high entropy attitude toward snakes, while both a snake enthusiast and phobic have low entropy attitudes toward snakes. The Gibbs entropy therefore measures the inherent stability of a system, which leads to the following proposition:

Proposition I.2: The Gibbs entropy of the attitude network reflects the attitude's stability.

From Proposition I.1, it follows that the natural state of an attitude is neutral or ambivalent and that consistent attitudes should be rare. However, this is clearly not the case, as even while individuals are often exposed to ambiguous information, they often arrive at consistent representations of the information (e.g., Holyoak & Simon, 1999; Simon & Spiller, 2016). So, why are attitudes often consistent whereas playing slot machines generally results in losing your money? The answer to this is that attitude elements are not independent of each other (to be explained below) and because of this dependency attitudes can assume low entropy macrostates. However, for a system to remain in a low entropy state (i.e., low Gibbs entropy) force has to be put on this system and it is our view that one of the main functions of focusing our attention on (or thinking about) an attitude object is to put such force on the attitude system and obtain (or maintain) a consistent attitude that is low in entropy.

Entropy reduction is a crucial aspect of life because a key characteristic of any living organism is that it must maintain order in its own system (Schrödinger, 1944). According to Kauffman (1993) the ability to reduce entropy is the most important selection criterion for evolution. This implies that the ability to reduce entropy is one of the central hallmarks of any living organism. We think that a similar argument can be made for the human mind, so that one of the central objectives of the human mind is to reduce its entropy (cf., Hirsh, Mar, & Peterson, 2012). It is straightforward that only attitudes low in entropy fulfill the functions typically associated with attitudes, such as to organize knowledge, increase utility, and express values (Katz, 1960; Smith, Bruner, & White, 1956). All these functions require attitudes to be in predictable, stable, and consistent states and therefore attitudes are much more likely to fulfill their functions when they are low in entropy (e.g., only a low entropy attitude toward snakes can clearly imply that you should run when you are near one). Linking the need for entropy reduction to cognitive consistency also echoes the fundamental and widespread assumption in research on attitudes that individuals have an inherent preference for cognitive consistency (e.g., Festinger, 1957; Gawronski & Strack, 2012; Heider, 1946, 1958; Monroe & Read, 2008; Shultz & Lepper, 1996).

6.2.3 The Causal Attitude (CAN) Model

To formalize the ideas presented here, we build on the Causal Attitude Network (CAN) model (Dalege et al., 2016), which treats attitude elements as nodes in a network that are connected by pairwise interactions. The complexity of the attitudinal representation is reflected by the size of the network (i.e., number of nodes). The CAN model is based on psychometric network models (e.g., Cramer et al., 2010; van der Maas et al., 2006) and on constraint-satisfaction models of attitudes (e.g., Kunda & Thagard, 1996; Monroe & Read, 2008; Shultz & Lepper, 1996). The central assumption of the CAN model is that dynamics of attitude networks can be described in an idealized way by the Ising (1925) model, which originated from statistical mechanics. While the Ising model is an extremely parsimonious model, its behavior is exceptionally rich. Due to these qualities, the Ising model has been applied to many different fields of research, such as magnetization (e.g., Ising, 1925), kinetic energy (e.g., Fredrickson & Andersen, 1984), predator-prey dynamics (e.g., Kim, Liu, Um, & Lee, 2005) neuroscience (e.g., Fraiman, Balenzuela, Foss, & Chialvo, 2009), clinical psychology (e.g., Cramer et al., 2016), and population dynamics (e.g., Galam, Gefen, & Shapir, 1982).

The Ising model describes the dynamics of networks by utilizing the fact that systems strive towards low energy configurations (see Figure 6.2 for an illustration of a simple Ising model). The energy of a configuration is determined by two classes of parameters. The first class constitutes the thresholds of the nodes, which determine the disposition of a given node to be ”on” or ”off” (denoted as τ_i). A node with a positive (negative) threshold requires less energy when it is ”on” (”off”). In the original Ising model, thresholds represent the external field that influences the spins of the magnet. Similarly, in attitude networks thresholds represent external information regarding the attitude object. These thresholds therefore represent the disposition of a given attitude element to be endorsed or not. A positive threshold represents a disposition of a given node to be ”on” (e.g., a positive thresholds of judging snakes as dangerous indicates that one is inclined to judge snakes as dangerous holding all other information in the attitude network constant). A negative threshold represents a disposition of a given node to be ”off” (e.g., a negative threshold of judging snakes as beautiful indicates that one is inclined to judge snakes as *not* beautiful). The magnitude of thresholds can also vary and the higher the magnitude, the stronger the disposition of the node to be ”on” or ”off”. In the Ising model shown in Figure 6.2, two nodes have the disposition to be ”on” (indicated by green thresholds) and two nodes have the disposition to be ”off” (indicated by red thresholds).

The second class of parameters constitutes weights of edges between nodes, representing the strength of interaction between nodes (denoted as ω_i). Two nodes that have positive weights between them require less (more) energy when they assume the same (different) state, representing preference for consistency. A positive weight represents an exhibitory interaction (e.g., feeling afraid of snakes because you also judge them as dangerous) and a negative weight represents an inhibitory interaction (e.g., not judging snakes as beautiful because you judge them as dangerous). The magnitude of weights can also vary and the higher the magnitude, the stronger the interaction. The CAN model assumes that weights between attitude elements generally arise based on inferences that support evaluative consistency. In the Ising model shown in Figure 6.2 all nodes are positively connected (indicated by green edges). This Ising model thus represents a simple attitude network consisting of, for example, four positive beliefs (e.g., believing that snakes maintain ecological order and are safe, beautiful, and smooth). Note that in the current chapter we focus on the situation in which edges between attitude elements are already present. How we can model the development of edges in attitude networks is currently investigated in our laboratory. The starting point for this investigation is to combine the AE framework with connectionist models of attitudes, which assume that Hebbian learning underlies development of attitudinal structures (e.g., Monroe & Read, 2008).

Thresholds and weights determine a given configuration’s energy (de-

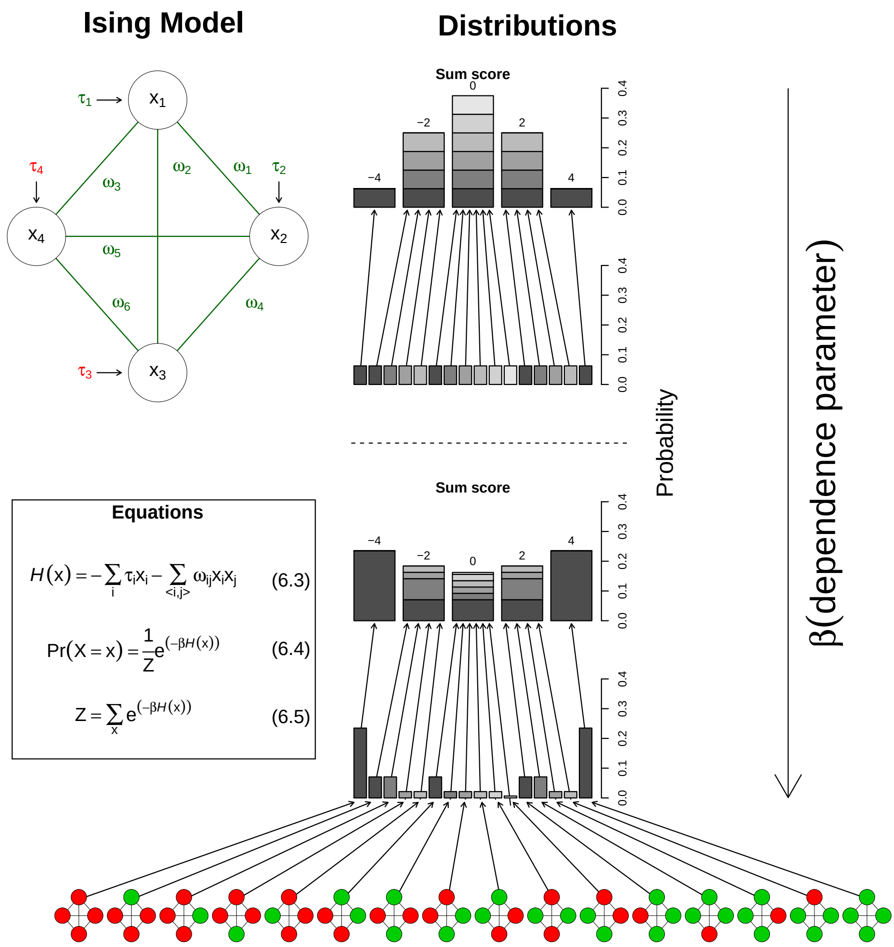


Figure 6.2: Illustration of the Ising model. In (6.3) H represents the Hamiltonian energy of the configuration of k distinct nodes $1,...,i,\ j,...k$, that engage in pairwise interactions and the variables x_i and x_j represent the states (-1, +1) of nodes i and j, respectively. The parameter τ_i represents the threshold of node i and the parameter ω_{ij} represents the interaction weight between nodes i and j. In (6.4) $Pr(X = x)$ represents the probability of a given network configuration and β represents the dependence parameter of the Ising model. In (6.5) Z represents the standardization factor, which ensures that the probabilities add up to 1. The Distributions part of the figure shows the probability distributions of two Ising models for the sum scores of the nodes (upper distributions) and the individual configurations (lower distributions). The bottom of the figure shows all possible states of the four-node network, with green (red) nodes indicating that the node is "on" ("off").

noted by H). It is our view that, in contrast to the physical application of the Ising model, energy does not reflect an existing physical property. The reason that calculation of energy is needed is that it enables the mental system to arrive at a low-entropy state by evaluating locally which elements need to be changed. By evaluating several attitude elements in turn, the mental system is able to create a global low-entropy state without evaluating the global state directly (which would probably be too complex from a computational point of view). This leads to the following proposition:

Proposition II: Energy of the attitudinal representation serves as a local processing possibility to evaluate the global Boltzmann entropy of an attitude. Attitude elements are likely to change when the opposite state has lower energy.

The extent to which a configuration's energy results in the configuration with lower energy being more likely than a configuration with higher energy depends on the dependence parameter (representing temperature in the original Ising model). The higher the dependence parameter, the more the probability of a configuration depends on its energy. Because of this, the dependence parameter directly scales the Gibbs entropy of a given Ising model (e.g., Kindermann & Snell, 1980) – implying that increasing the dependence parameter results in attitude networks being more ordered and stable. For example, the Ising model with dependence at 0 at the top of Figure 6.2 also has maximum Gibbs entropy, because all configurations are equally likely. In contrast, the Ising model with high dependence at the bottom of Figure 6.2 has lower Gibbs entropy, because the completely consistent configurations are much more likely than the inconsistent configurations. A system low in Gibbs entropy thus creates the possibility of macrostates having low Boltzmann entropy, but as long as the system is not at minimum Gibbs entropy, macrostates with high Boltzmann entropy are still possible.

The probability equation allows us to calculate the distribution of configurations we would expect if we measure an infinite number of individuals holding an attitude that can be described by a given Ising model. For describing the dynamics of a given individual's attitude we can utilize time dependent dynamics called Glauber dynamics (Glauber, 1963). The basic workings of Glauber dynamics on Ising models are that at each iteration we (i) calculate the energy of the current configuration, (ii) pick a random node and calculate the energy of this neighboring configuration when this node is "flipped" (e.g., when this node changes from "on" to "off"), (iii) determine the probability of the node actually flipping by us-

Detailed Balance	$\frac{\Pr(x_i)}{\Pr(x_j)} = e^{(\Delta E_{x_i - x_j}\beta)}$	(6.6)
Flip Probability Equation	$\Pr(x_i \rightarrow -x_j) = 1/(1 + e^{(\Delta E_{x_i - x_j}\beta)})$	(6.7)

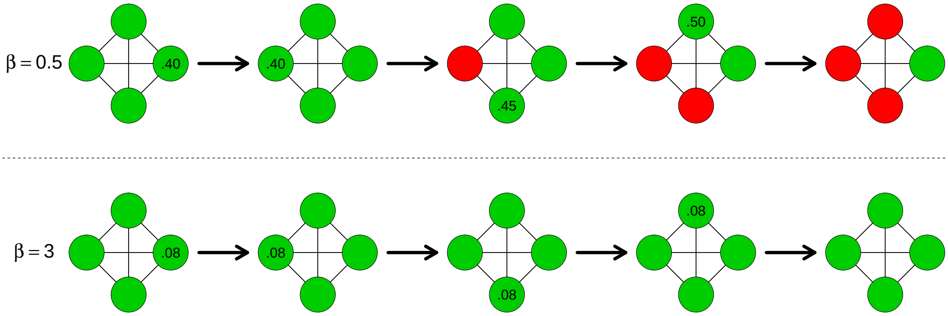


Figure 6.3: Glauber dynamics for two four-node networks under different dependence parameters. In (6.6) and (6.7) x_i and x_j represent the current state of a given node and its opposite state, respectively. (6.6) represents the probability of a node remaining in its current state relative to the probability that it will flip. Each network represents one iteration and the node with a given probability represents a node that was randomly picked to be flipped with the given probability.

ing the difference in energy, and (iv) flip the node with this probability (see Figure 6.3 for an illustration and equation). For attitude dynamics, this implies that increasing an attitude's consistency can be described by such dynamics. For example, if one believes that snakes are safe, while one also feels scared of them and always screams when one sees a snake, the probability that one changes her belief that snakes are safe is high. In the simulations we describe later, we make use of Glauber dynamics when we model individual-level dynamics.

Figure 6.3 illustrates the reason why the dependence parameter scales the Gibbs entropy of an Ising model. In the network with the dependence parameter at 0.5, the thresholds and weights have little influence on the network's dynamics and the network behaves essentially randomly. This situation therefore represents an attitude that is unstable and in which the different attitude elements are held with low certainty (e.g., the attitude of a person, who does not care at all about snakes). With increases in the dependence parameter, the probability of the network configuration becomes more and more dependent on the thresholds and the weights of the network. In the network with the dependence parameter at 3, the thresholds and weights have strong influence on the network's dynamics and the network behaves in accordance with these parameters. This situation therefore represents an attitude that is stable and in which the different attitude elements are held with high certainty (e.g., the attitude of a snake phobic or enthusiast). This underscores the necessity of attitude elements being dependent on each other to reduce attitudinal entropy. In this chapter, we argue that the dependence parameter in the Ising model constitutes a formalized representation of the effect of directing attention and thinking about attitude objects, leading to the third proposition of the AE framework:

Proposition III: Focusing attention on the attitude object and thinking about the attitude object reduces the Gibbs entropy of attitudes by increasing the attitude network's dependence parameter. The higher the dependence parameter the stronger the correspondence between energy and probability of a given attitude network's configuration.

Based on our argument that individuals are motivated to reduce attitudinal entropy, we expect that a high level of attitudinal entropy causes psychological discomfort. However, because we assume that entropy of attitude networks cannot be directly evaluated, this influence is indirect. We expect that both measures of attitudinal entropy translate into psychological discomfort through proxies, which are easier to evaluate for the mental system. This leads to the following implications:

Implication I: High Boltzmann entropy in combination with a high dependence parameter indirectly leads to psychological discomfort. The Boltzmann entropy is indirectly evaluated by the difference in energy of the current and neighboring configurations (i.e., configurations for which only one attitude element has to be flipped).

Implication II: High Gibbs entropy in combination with a high dependence parameter indirectly leads to psychological discomfort. The Gibbs entropy is indirectly evaluated by the temporal stability of the attitude.

6.3 Levels of Attitudinal Entropy Reduction

In this section, we discuss different levels of attitudinal entropy reduction and research supporting these different levels. Note that these levels do not represent distinctive categories, but are assumed to lie on a dimension from weak entropy reduction to high entropy reduction (just as the dependence parameter in the Ising model is also a continuous variable). It is our view that thinking about an attitude object, or more generally, paying attention to an attitude object has the default effect of slightly increasing the dependency of the attitude network; as such, simply focusing attention on the attitude object represents the most basic level of dependency of the attitude network. Such a situation, for example, arises when an individual observes an attitude object. The dependence parameter increases when the individual is prompted to think about the attitude object, which would, for example, be the case when the individual responds to a questionnaire about an attitude object.

Increased levels of attitudinal entropy reduction may arise when individuals are for some reason prompted to think more elaborately about an attitude object and dependency of the attitude network is further increased when motivational factors come into play, representing intermediate levels of attitudinal entropy reduction. Examples of factors moderately increasing motivation to reduce attitudinal entropy are situations in which individuals are committed to an evaluation or in which they have to make a relatively unimportant decision.

Even more enhanced levels of attitudinal entropy reduction arise when individuals attach personal importance to their attitudes. Attitude importance is a widely researched topic and is a key determinant of attitude strength (Howe & Krosnick, 2017). Factors increasing attitude importance are the attitude's relevance to self-interests (e.g., attitude's relevance to important decisions), to personal values, and to social identification (Boninger et al., 1995). Crucially, attitude importance is strongly related to how much attention individuals devote to an attitude object (Krosnick et al., 1993), making it likely that indeed attitudes high in personal importance represent attitude networks with the highest dependence and therefore also the lowest entropy attitudes. Furthermore, as we discuss in a later section, strong attitudes show exactly the dynamics that would be expected from high dependence attitude networks.

To summarize, lower levels of attitudinal entropy reduction arise when individuals pay some attention to the attitude object or briefly think about it. Intermediate levels represent situations, in which an individual is for some reason prompted to think about the attitude object in more detail (e.g., when an argument regarding the attitude object has to be evaluated) and when individuals have to base a decision on their attitude network or are committed to an evaluation. The final levels of attitudinal entropy reduction arise when an individual attaches high personal importance to an attitude object. In the remainder of this section, we show that several central findings in the attitude literature follow from the entropy reducing function of attention and thought.

6.3.1 Implicit Measures are more Likely to Tap Attitudes in High Entropy States

The dependence parameter of attitude networks increases when attention is directed at the attitude object, which implies that the measurement of attitudes influences the dependence parameter. Implicit measures of attitudes, such as the Implicit Association Test (IAT; Greenwald et al., 1998) and the Affective Misattribution Task (AMP; B. K. Payne et al., 2005), limit attention directed at the attitude object by measuring attitudes without directly asking individuals to introspect. These measures are therefore more likely to tap attitudes in high entropy states than explicit measures. Attitudes in high entropy states are less internally consistent than attitudes in low entropy states, which might contribute to the fact that implicit measures generally show both poor internal reliability and test-retest reliability (e.g., Bar-Anan & Nosek, 2014; Gawronski, Morrison, Phills, & Galdi, 2017; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). In contrast to the low individual temporal stability of scores on implicit measures of attitudes, mean effects on implicit measures are substantially more robust (B. K. Payne, Vuletich, & Lundberg, 2017). For example, children show similar scores on the IAT as adults (Baron & Banaji, 2006) and the IAT predicts behavior much better on a global level (e.g., police shootings of Blacks is strongly associated with prejudice assessed with the IAT on a regional level; Hehman, Flake, & Calanchini, 2018) than on an individual level, which is generally rather low (e.g., Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013). A recent review identified these patterns as important puzzles in the literature on implicit measures of attitudes (B. K. Payne et al., 2017). With the following simulations we show that these puzzles can be straightforwardly solved by assuming that implicit measures are more likely to tap attitudes in high entropy (instable) states.

Simulation 1a: Implicit Measures – low temporal stability, stable means. For this simulation, we investigated Glauber dynamics on a fully connected ten-node network with all edges set to .1. We varied the thresholds uniformly from -.2 to .7 with steps of .1, resulting in 10 different sets of thresholds (we chose these thresholds, so that the means differ from 0 and that there is sufficient variance for correlations to be meaningful). The dependence parameter was set to .5 (representing that only some attention is directed at the attitude object) and the network was randomly initialized. We simulated 100 individuals for each set of thresholds, resulting in the total number of 1,000 individuals. For each individual we simulated 1,000 iterations. After 500 and 1,000 iterations, respectively, we measured the sum score of the nodes. The resulting scores at the first measurement and at the second measurement were only weakly correlated, $r = .24$, $p < .001$. The means of the first measurement ($M = 1.91$) and the second measurement ($M = 2.04$), in contrast, were virtually identical, t (999) = 0.71, $p = .479$, although the standard deviations at both the first measurement ($\sigma = 4.55$) and the second measurement were substantial ($\sigma = 4.45$). Networks under a low dependence parameter thus show low individual temporal stability, but stable means, which precisely matches the known behavior of implicit attitude measures.

Simulation 1b: Implicit measures – low behavior prediction on individual level, high behavior prediction on group level. For this simulation, we used mostly the same setup as in Simulation 1a with the following adjustments. First, we added an eleventh node that represented behavior. This node was also connected with weights of .1 to all other nodes, but always had a threshold of 0 (so that all systematic variation in this node is caused by its connection to other nodes). Second, we ran a total number of 10,000 individuals to be able to create groups of sufficiently large size. Third, we created ten groups that differed in their mean thresholds (see Appendix E). The correlation between the first ten nodes and the 'behavior' node on an individual level was relatively weak, $r = .23$, $p < .001$. In contrast, the correlation on the group level was very strong, $r = .80$, $p = .006$. These patterns fit the finding that individual level correlations between implicit measures of attitudes and behavior are relatively low and that group level correlations are considerably stronger.

The implication of the AE framework that implicit measures are more likely to tap attitudes in high entropy states than explicit measures has fundamental implications for the research on implicit measures of attitudes. While researchers in this domain have long acknowledged the fact that implicit measures show low internal consistency, they have generally interpreted this as a measurement problem (e.g., Fazio & Olson, 2003b; Gawronski et al., 2007; LeBel & Paunonen, 2011; Nosek & Banaji, 2001). However, the AE framework implies that the construct measured by implicit measures is *itself* more internally inconsistent than the construct measured by explicit measures because the former by their very nature direct less attention towards the attitude object than the latter. One consequence of this is that often the only way to make implicit measures more reliable is to make them more explicit – leading to the counterintuitive conclusion that a valid measurement of attitudes (or any system for that matter) in high entropy states *must* be unreliable.

Prediction 1a: Manipulating the dependency in attitude networks (e.g., letting individuals think for some time about the attitude object) is expected to increase internal consistency and stability of implicit measures.

Prediction 1b: Scores on implicit measures assessing attitudes, individuals regularly think about, are expected to have higher internal consistency and stability than scores on implicit measures assessing attitudes, individuals think only infrequently about.

Prediction 1c: Implicit and explicit measures should show the lowest convergence when the dependence of the attitude network is generally low.

6.3.2 The Mere Thought Effect as an Initial Level of Heightened Attitudinal Entropy Reduction

The mere thought effect on attitude polarization refers to the classic finding that briefly thinking about an attitude object without receiving external information results in more extreme evaluation of the attitude object (e.g., Tesser, 1978; Tesser & Conlee, 1975). Based on several studies on the mere thought effect, Tesser, Martin, and Mendolia (1995) argued that (1) sufficiently complex cognitive schemas (defined as the number of dimensions an attitude object is rated on) are necessary (Tesser & Leone, 1977) and (2) the evaluative dimensions, on which the cognitive schema is based, need to be sufficiently interdependent for the mere thought effect to manifest itself (Millar & Tesser, 1986). In the following simulations, we show that the mere thought effect and its moderators naturally follow from the AE framework.

Simulation 2a: Basic mere thought effect. We calculated the probabilities of the sum scores of a fully connected ten-node network with the dependence parameter set to either 1 (representing merely asking individuals about their attitudes) or 1.5 (representing mere thought). All edge weights were set to .1 and all thresholds were set to 0. As can be seen in Figure 6.4a, increasing the dependence parameter leads to an increase in extreme sum scores, mimicking the basic mere thought effect.

Simulation 2b: Network size as formalization of a complex cognitive schema As stated in the section on the CAN model, the size of networks reflects the complexity of the cognitive schema of the attitude object. We therefore expect that a network with few nodes will not show a strong increase in extreme sum scores when the dependence parameter is increased. To investigate this, we adapted Simulation 2a by decreasing the number of nodes to 4. As can be seen in Figure 6.4b, increasing the dependence parameter for such a small-size network does not lead to a substantial increase in extreme sum scores, mimicking the finding that

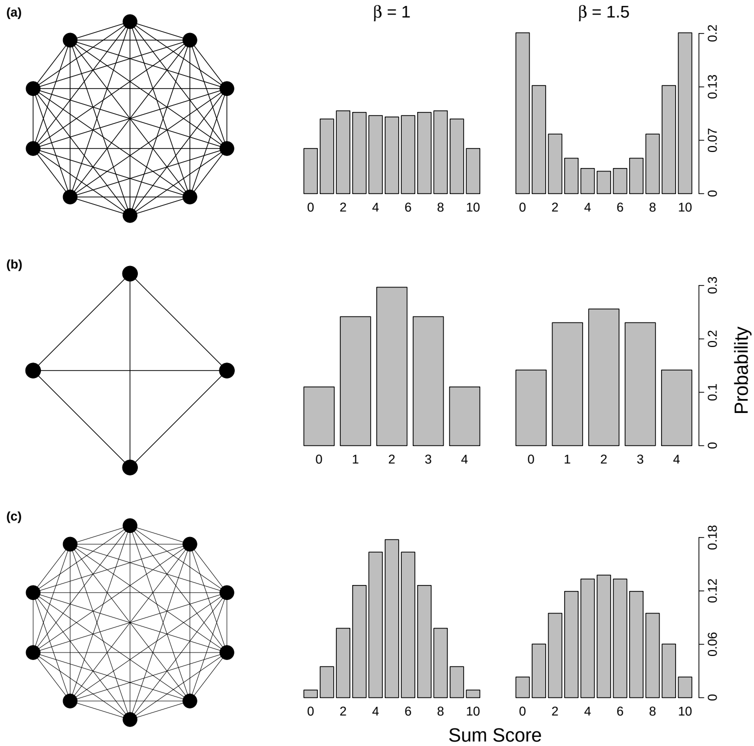


Figure 6.4: The mere thought effect on polarization modeled by the AE framework. Implied distributions of sum scores are shown under different dependence parameters for (a) a ten-node network with all edge weights equal to .1, (b) a four-node network with all edge weights equal to .1, and (c) a ten-node network with edge weights equal to .05.

complex cognitive schemas are necessary for the mere thought effect to manifest itself.

Simulation 2c: Magnitude of edge weights as a formalization of dependence between evaluative dimensions. Edge weights in Ising networks are a straightforward formalization of interdependence between evaluative dimensions. To investigate whether decreasing edge weights leads to lower increase in extreme scores, we adapted Simulation 2a by setting all edge weights to .05. As can be seen in Figure 4c, increasing the dependence parameter for such a weakly connected network does not lead to a substantial increase in extreme sum scores, mimicking the finding that interdependence of complex schemas is necessary for the mere thought effect to manifest itself.

Further support for the proposition that merely thinking about an attitude object represents an intermediate level of attitudinal entropy reduction comes from research on the coherence effect in judgment and decision-making (e.g., Holyoak & Simon, 1999; Simon, Snow, & Read, 2004; Simon, Pham, Le, & Holyoak, 2001; Simon, Krawczyk, & Holyoak, 2004; Simon et al., 2015). The coherence effect represents the general finding that when individuals are presented with ambiguous information about a given scenario (e.g., a legal case), individuals interpret this information in such a way that it allows for a coherent judgment or decision about the scenario. Interestingly, such coherence shifts are also observed for dependency between emotions and beliefs regarding an attitude object (Simon et al., 2015) and in the complete absence of making a decision (Simon et al., 2001). However, there are some indications that having to make a decision heightens the coherence shift effect (Simon et al., 2001; Simon, Krawczyk, & Holyoak, 2004).

Prediction 2: Sizes of edge weights and size of the attitude network predict the strength of the mere thought effect.

Prediction 3: Because the AE framework assumes that increasing dependency of attitude networks is a continuous process, the AE framework predicts that also an opposite mere thought effect exists, in the sense that when individuals are asked to very quickly answer attitude questions, attitudes are expected to be *less* polarized than when individuals are given more time to answer the questions. Note that the AE framework predicts that this would constitute a small effect.

6.3.3 Attitude Strength

The highest levels of attitudinal entropy reduction have implications for attitude strength. The macro-behavior of Ising networks is governed by the dependence of the network and can be described by the cusp catastrophe model (Sitnov, Sharma, Papadopoulos, & Vassiliadis, 2001). The cusp catastrophe model describes sudden vs. smooth changes in a variable depending on two control variables, referred to as the normal and splitting variable, respectively (Gilmore, 1981; Thom, 1972; Zeeman, 1976). Depending on the value of the splitting factor, the influence of the normal variable on the dependent variable is either gradual or discrete, implying that the so-called bifurcation area, in which sudden transitions happen in the dependent variable, is larger when the splitting factor is high (see Figure 6.5). As an illustration, take the freezing of water. In this case, temperature represents the normal variable and pressure represents the splitting variable – under low pressure water freezes and melts at the same temperature, while under high pressure frozen water melts at a higher temperature than when liquid water freezes (and the other way around).

In Ising networks the average of the thresholds functions as normal control variable, the dependence of the network as splitting variable, and the macro-behavior of the network as dependent variable (Sitnov et al., 2001). Because of this, networks high in dependency are stable, ordered, and change happens suddenly, while networks low in dependency are fluctuating, random, and change happens gradually. These observations link the AE framework to the catastrophe model of attitudes, which assumes that attitude change can be described by the cusp catastrophe model (Flay, 1978; Latané & Nowak, 1994; Zeeman, 1976). In the catastrophe model of attitudes, valenced information functions as the normal variable, attitude involvement or attitude importance functions as the splitting variable, and the global evaluation functions as the dependent variable (see Figure 6.5). Several studies support the catastrophe model of attitudes by showing that important attitudes are more extreme than unimportant attitudes (e.g., Latané & Nowak, 1994; Liu & Latané, 1998) and by directly fitting the catastrophe model to data on attitudes (van der Maas et al., 2003). The CAN model can easily integrate the catastrophe model of attitudes and also provides a micro-level explanation of the postulates of the catastrophe model (Dalege et al., 2016). Thresholds in the CAN model directly relate to the valenced information a person receives regarding an attitude object and the macro-behavior of an attitude is strongly related to global evaluations of the attitude object. These similarities lead to the conclusion that important attitudes are based on attitude networks high in dependence.

Linking attitude importance to the dependence of attitude networks also has broader implications for attitude strength. As attitude importance is a central determinant of attitude strength (Howe & Krosnick, 2017), it becomes likely that strong attitudes represent high dependence attitude networks. Indeed, the dynamics of strong attitudes are highly similar to the dynamics of strongly connected networks (Dalege et al., 2016), which in turn are similar to the behavior of low dependence networks. Similar to strong attitudes (Krosnick & Petty, 1995), high dependence networks are more stable and resistant (Kindermann & Snell, 1980). Increasing the dependence of attitude networks likely results in information being processed in accordance with the attitude, which represents another central feature of attitude strength.

Biased information processing is related to the phenomenon of hysteresis in the cusp catastrophe model. Hysteresis implies that the point at which a system moves to the opposite state depends on the direction of change (just as is the case for the melting and freezing of water under

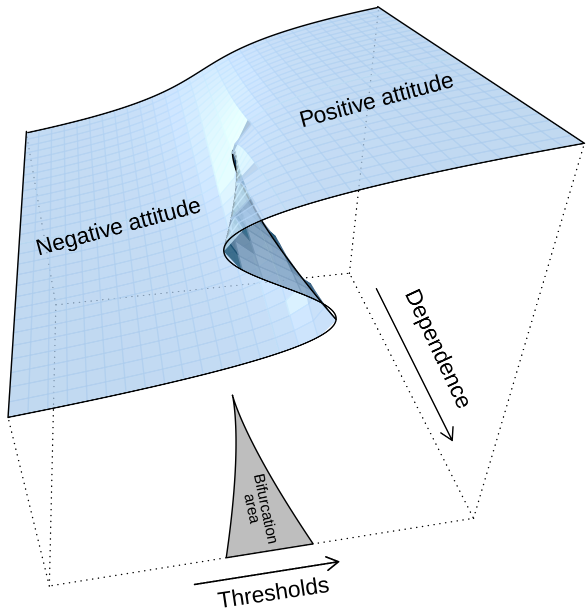


Figure 6.5: The cusp catastrophe model from the perspective of the AE framework.

high pressure). The strength of the hysteresis effect in the cusp catastrophe model depends on the splitting variable – implying that attitude networks under high dependence should show strong hysteresis effects (i.e., the bifurcation area becomes broader). Changing such an attitude thus requires a disproportionate amount of persuasion compared to the amount of information the individual already received. In such a situation it would probably be more effective to first reduce the dependence parameter of the attitude network, so that the individual is more 'open' to change.

Attitude-behavior consistency, which represents the final central feature of attitude strength, is also more likely in high dependence attitude networks, because attitude elements are more dependent on each other. As the CAN model treats behavior as part of the attitude network, increasing dependence of attitude networks also increases the dependence of behavior on beliefs and feelings regarding the attitude object, and vice versa, implying higher attitude-behavior consistency (Dalege, Borsboom, van Harreveld, Waldorp, & van der Maas, 2017). This hypothesis is supported by findings indicating that attitude-behavior consistency depends on the stability of attitudes (Glasman & Albarracín, 2006), which represents a proxy of the dependence of the attitude network.

Prediction 4: The mere thought effect extends to the stability and resistance of attitudes. Thinking briefly about attitude objects is predicted to (temporally) increase the stability and resistance of attitudes.

Prediction 5: Persuasion is more effective for strong attitudes when the dependence parameter is lowered (e.g., by reducing attention directed at the attitude object) before the persuasion is employed. The reason for this is that lowering the dependence parameter reduces the hysteresis effect.

Prediction 6: Whether an attitude changes continuously or discretely depends on the dependence parameter of the attitude network.

Prediction 7: Reducing the dependence parameter of networks (e.g., by limiting cognitive capacity) results in less stable and less resistant attitudes.

6.3.4 Heuristic-Based vs. Argument-Based Persuasion as Global vs. Specific Threshold Changes

The Elaboration Likelihood Model (Petty & Cacioppo, 1986) and the Heuristic Systematic Model (Chaiken et al., 1989) are two hallmark dual process theories assuming that persuasion can be accomplished via two routes – one in which individuals change their attitudes based on heuristic cues (e.g., whether the source of the message is an expert) and one in which individuals change their attitudes based on a deeper processing of the quality of the persuasive arguments. Several studies have supported this idea and showed that individuals low in involvement are more likely to change their attitudes according to heuristic cues, while individuals high in involvement are more likely to change their attitudes according to argument quality (e.g., Petty & Cacioppo, 1984; Petty, Cacioppo, & Goldman, 1981; Petty, Cacioppo, & Schumann, 1983). From the perspective of the AE framework, heuristic-based persuasion represents a moderate global change in the attitude network's thresholds (i.e., change in the magnetic field in the language of the original Ising model), while argument-based persuasion represents a strong change of few specific thresholds – implying that moderate global change is more influential under low dependence, while strong specific change is more influential under high dependence. We tested this hypothesis in the following simulation.

Simulation 3: Global vs. specific threshold changes. For this simulation, we again used a fully connected ten-node network with all edge weights set to .1. We investigated Glauber dynamics of this network using 1,000 iterations. In the first 500 iterations, all simulated individuals' thresholds were set to .2 (thus representing a positive initial attitude) and the network was randomly initialized. In the second 500 iterations, (a) thresholds either remained at .2 (representing the no heuristic cue/weak arguments condition), (b) all thresholds changed to -.12 (representing the heuristic cue/weak arguments condition), (c) the first four thresholds changed to -.6, while the other thresholds remained at .2 (representing the no heuristic cue/strong arguments condition), or (d) the first four thresholds changed to -.72 and the other thresholds changed to -.12 (representing the heuristic cue/strong arguments condition). The thresholds were chosen in this way, so that the mean change in the heuristic cue/weak arguments condition and the no heuristic cue/strong arguments condition was equal. Half of the simulated individuals' βs was set to 1 (representing low involvement) and the other half of the simulated individuals' βs was set to 3 (representing high involvement). We simulated 100 individuals for each experimental cell, resulting in 600 simulated individuals in total. The results showed a pattern reminiscent of the typical findings in the heuristic-based vs. argument-based persuasion literature (e.g., Petty et al., 1981). The three-way interaction on the sum score of the attitude elements at the $1,000^{th}$ iteration was significant, F (1, 792) = 16.40, p = .001, η_p^2 = .02, and the pattern of the results was in line with the hypotheses, see Figure 6.6. Conceptualizing heuristic cues as global moderate threshold change and strong arguments as specific strong thresholds change thus explains the basic result in the heuristic-versus argument-based persuasion literature.[3]

Prediction 8a: Sufficiently strong heuristic cues lead to attitude change under both low and high involvement.

Prediction 8b: A large number of strong arguments lead to attitude change under both low and high involvement.

Prediction 9: As can be seen in Figure 6b, specific threshold change also affected networks with low β to a meaningful extent. This leads to the prediction that, given sufficient power (e.g., in a meta-analysis), an effect of strong arguments should also be detected under low involvement.

[3]We want to emphasize that while in all the other simulations presented here the findings are highly robust to changes in parameters, for the current simulation specific parameters had to be chosen to find the reported pattern of the results (e.g., when global threshold changes are chosen that are too high or when too many specific thresholds are targeted, differences between the β conditions become less meaningful). The AE framework therefore predicts that the effect of argument vs. persuasion-based persuasion is limited to a specific range of stimuli (i.e., not too strong heuristic cues or not too many strong arguments).

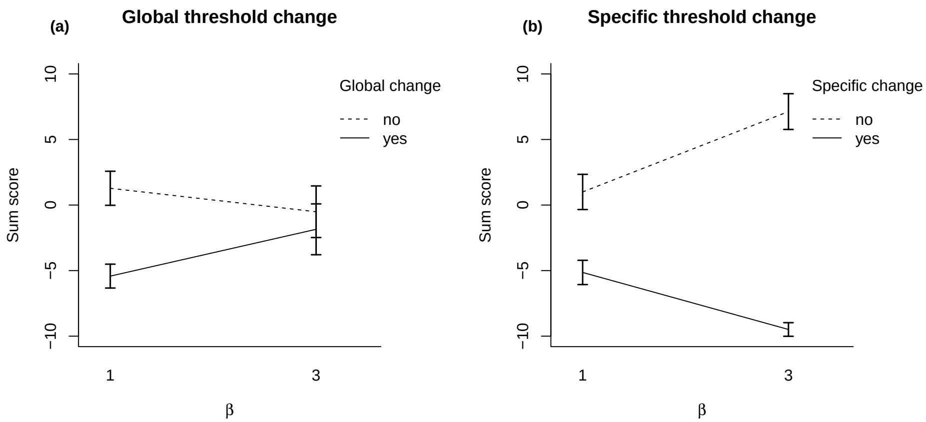


Figure 6.6: Effects of global (a) and specific threshold change (b) under moderate and high β. Error bars represent +- 2 standard deviations around the mean.

6.4 Aversiveness of Attitudinal Entropy: Ambivalence and Cognitive Dissonance from the Perspective of the AE Framework

The first proposition of the AE framework holds that inconsistency of an attitude is attitudinal entropy. Research on ambivalence and cognitive dissonance underscores that consistency is a fundamental human need and violations of this need cause psychological discomfort. Cognitive dissonance refers to aversive feelings caused by incongruent beliefs and behaviors vis-á-vis an attitude object, with most research on cognitive dissonance focusing on the effects of carrying out a behavior inconsistent with the beliefs an individual holds. A crucial distinction in the research on ambivalence is that between potential (or objective) and felt (or subjective) ambivalence (e.g., Newby-Clark et al., 2002; Priester & Petty, 1996; van Harreveld, van der Pligt, & de Liver, 2009).[4] Potential ambivalence refers to the number of incongruent attitude elements and felt ambivalence refers to the aversive feelings caused by these incongruent attitude elements. Crucially, the distinction between potential and felt ambivalence is made, because potential ambivalence *can*, but *not necessarily does* result in felt ambivalence. In other words: ambivalence can, but does not

[4] We use the terms potential and felt ambivalence throughout the chapter, because these terms fit our framework better than the recently more commonly used terms objective (or structural) and subjective ambivalence.

have to be, unpleasant. From the perspective of the AE framework, the question when potential ambivalence results in felt ambivalence becomes the question when attitudinal entropy results in psychological discomfort. In this section we discuss two possibilities how the mental system indirectly evaluates attitudinal entropy and under which circumstances this results in psychological discomfort. This discussion is based on Implication I and II of the AE framework. Implication I holds that Boltzmann entropy is indirectly evaluated through the energies of neighboring configurations. In the following subsection we show that this implication integrates the Gradual Threshold (GT) model of ambivalence (Priester & Petty, 1996) into the AE framework. Second, Implication II holds that Gibbs entropy is indirectly evaluated by the temporal stability of the attitude network's configuration.

6.4.1 Boltzmann Entropy as Ambivalence

An influential account of how potential ambivalence translates into felt ambivalence is the Gradual Threshold (GT) model (Priester & Petty, 1996). This model assumes a curvilinear relation between the number of conflicting evaluations (treated here as attitude elements) and felt ambivalence, in which felt ambivalence increases less as the number of conflicting attitude elements increases (e.g., the difference between holding no conflicting attitude element and holding one conflicting attitude element is larger than the difference between holding three conflicting attitude elements and holding four conflicting attitude elements). The specific equation of the GT model is the following:

$$Ambivalence = 5(C+1)^p - (D+1)^{1/C}, \tag{6.8}$$

where C refers to conflicting attitude elements (i.e., attitude elements that are incongruent to the majority of attitude elements) and D refers to the number of dominant attitude elements (i.e., attitude elements that are consistent with the majority of attitude elements). p determines the power function and was estimated by Priester and Petty (1996) to lie somewhere between .4 and .5. While Priester and Petty (1996) explicitly stated that these specific values are exogenous to the GT model, most research based on the GT model uses a value that lies between .4 and .5 to calculate expected felt ambivalence scores (e.g., Clark, Wegener, & Fabrigar, 2008; Refling et al., 2013). Clearly, a better understanding of the power function would provide us with more knowledge on when and why potential ambivalence results in felt ambivalence (high p values would indicate an almost linear relation, while low p values would indicate a steep relation).

From the perspective of the AE framework, the number of conflicting attitude elements is given by the configuration of the attitude network. We therefore expect that felt ambivalence as modeled by the GT model indirectly reflects Boltzmann entropy of the configuration of the attitude network. As stated in Implication I of the AE framework, Boltzmann entropy is indirectly evaluated by the energy difference between the current configuration and its neighboring configurations and that the psychological discomfort caused by the energy difference is amplified by the dependence parameter. Based on this reasoning, we expect the dependence parameter to determine the steepness of the relation between potential and felt ambivalence.

Simulation 4: Felt ambivalence as boltzmann entropy. For this simulation we again used a fully connected ten-node network with all edge weights set to .1 and all thresholds set to 0. We first calculated the GT model's implied ambivalence scores for each of the possible configurations and varied the power function parameter between .3, .5, and .7.

Using Equation 6.6 we then calculated the mean preference of each node to remain in its current state for each configuration. We then averaged these scores for each configuration with the same number of conflicting attitude elements (e.g., for each configurations in which all but one attitude elements are in the positive state). We calculated the distribution of preference scores with the dependence parameter set to 1, 1.5, and 2.5. As can be seen in Figure 6.7, varying the dependence parameter has an analogous effect as varying the power function of the GT model. Based on this finding, the AE framework implies that the dependence parameter determines to what degree potential ambivalence translates into felt ambivalence.[5]

Based on the finding that preferences of nodes to remain in their current states are faster decelerating under a high dependence parameter, we expect that factors increasing the dependence parameter also increase felt ambivalence. This hypothesis is indirectly supported by the finding that having to make a decision increases felt ambivalence (e.g., Armitage & Arden, 2007; van Harreveld, Rutjens, et al., 2009; van Harreveld, van der Pligt, & de Liver, 2009). The AE framework holds that basing a decision on an attitude increases the dependence parameter of the attitude network, which results in lower (relative) preference of nodes to remain in their current state if the current configuration is ambivalent.

[5]Note that with increasing dependence parameter the preference scores of almost all configurations increase. It therefore seems likely that the preference scores are evaluated relative to the dependence parameter.

Prediction 10: Dependence of attitude networks moderates the relation between potential and felt ambivalence. The higher the dependence, the stronger the impact of the first incongruent attitude elements.

Prediction 11: The AE framework assumes that dependence is increased when a decision has to be made. After the decision is made, dependence drops again. This implies that before a decision is contemplated *and* after a decision, correspondence between potential and felt ambivalence is expected to be lower than while contemplating the decision. The same holds for the stability and resistance of the attitude.

6.4.2 Gibbs Entropy as Ambivalence

In our view, it is likely that felt ambivalence, if caused by low preference of nodes to remain in their current state, generally represents a situation-dependent process (e.g., having to make a decision based on an ambivalent attitude). In contrast, felt ambivalence caused by Gibbs entropy reflects a more chronic state of felt ambivalence. Implication II of the AE framework holds that the Gibbs entropy of an attitude network is indirectly evaluated by the stability of the attitude. Based on this implication, we expect that unstable attitude networks in combination with a high dependence parameter cause strong feelings of ambivalence. To investigate under which circumstances low stability and a high dependence parameter can co-occur, we set up the following simulation.

Simulation 5: Felt ambivalence as Gibbs entropy. For this simulation we again used a fully connected ten-node network with all edge weights set to .1. We varied the thresholds in the following way: (1) all thresholds were set to 0 (representing a situation in which the external information points in no direction), (2) half of the thresholds were set to .1 and the other half were set to -.1 (representing a situation in which an individual receives weak mixed information about an attitude object), (3) half of the thresholds were set to .5 and the other half were set to -.5 (representing a situation in which an individual receives strong mixed information about an attitude object), (4) all thresholds were set to .1 (representing a situation in which the external information points in a weak positive direction), all thresholds were set to .5 (representing a situation in which the external information points in a strong positive direction). Additionally, we varied the dependence parameter between 1, 1.5, and 2.5 (mirroring Simulation 4). For each combination of thresholds and dependence parameter, we simulated 100 individuals, resulting in the total number of 1,500 simulated individuals. For each individual, we simulated 500 iterations based on Glauber dynamics (the network was again initialized randomly). To

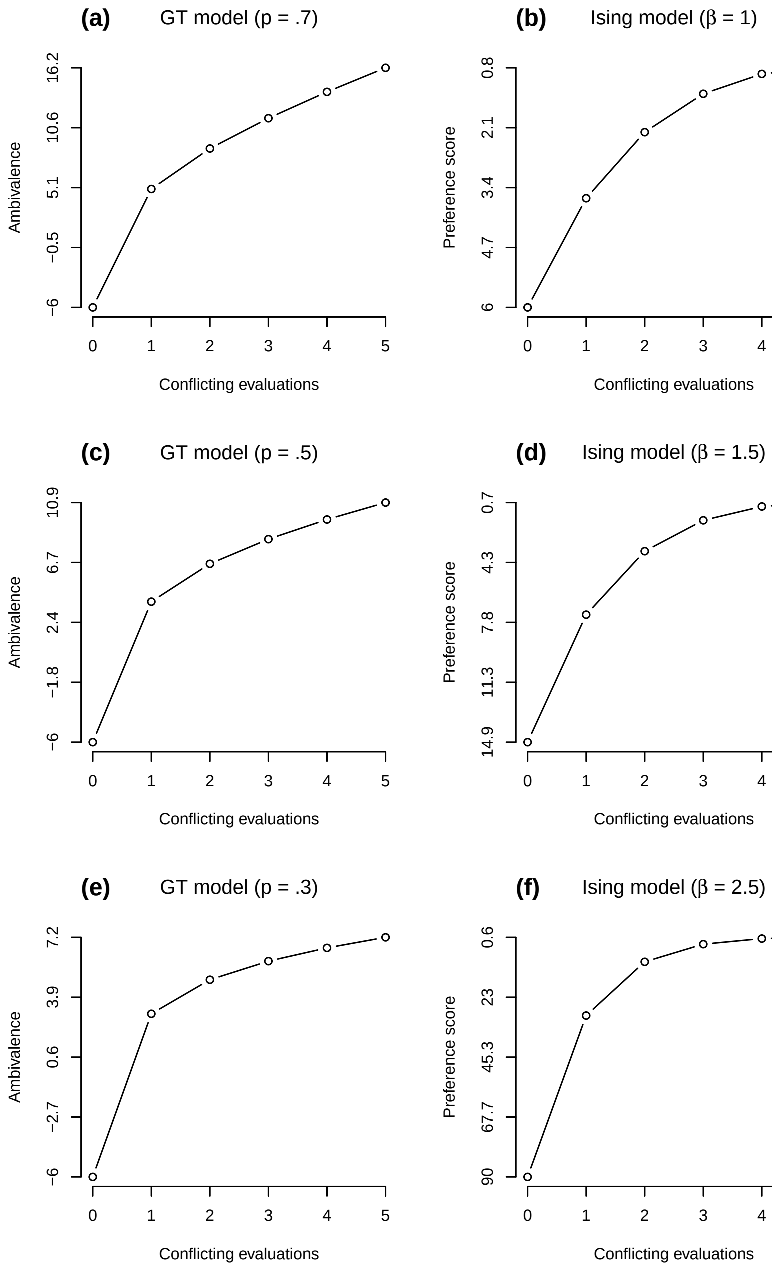


Figure 6.7: Similarity of ambivalence calculated by the GT model and mean preference scores of nodes in Ising networks.

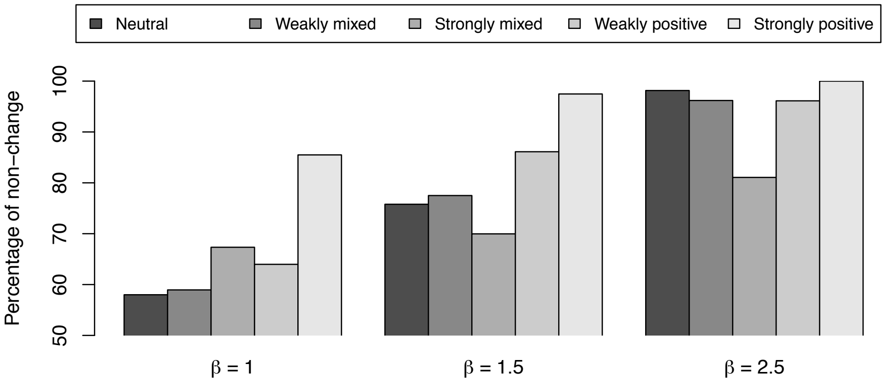


Figure 6.8: Stability of attitude networks based on different thresholds and dependence parameters.

evaluate the stability of the attitude network, we calculated the percentage of flipped states for the last 100 iterations.

The results indicated that for $\beta = 1$ only the highly positive thresholds resulted in a relatively stable attitude network (see Figure 6.8). For $\beta = 1.5$, attitude networks were more stable overall, with the strong positive thresholds resulting in almost perfect stability. The highly mixed thresholds networks remained relatively unstable. For $\beta = 2.5$, only highly mixed thresholds networks did not approach perfect stability. It is our view that such a situation results in the strongest feelings of ambivalence, because stability remains rather low while the dependence parameter is already at a high value.

Based on the results of the simulation, we conclude that high felt ambivalence arises when individuals receive highly mixed information. Felt ambivalence is then amplified by the motivation to reduce attitudinal entropy. Such a situation would arise when individuals hold important attitudes for which they receive mixed information; for instance, when individuals are disposed to a given evaluation (e.g., holding liberal values, because you work at a liberal university), while significant others endorse a different evaluation (e.g., having parents, who hold conservative values). Such a situation has been shown to cause strong feelings of ambivalence (Priester & Petty, 2001). Further support for the relation between stability of an attitude and feelings on ambivalence comes from the finding that ambivalent individuals show physical signs of instability (i.e., moving from one side to the other; Schneider et al., 2013).

Prediction 12: Highly-mixed information and high attitude importance result in strong felt ambivalence.

6.4.3 Cognitive Dissonance and Ambivalence Reflect Attitudinal Entropy

Apart from research on ambivalence, the implication that attitudinal entropy causes psychological discomfort is also relevant to research on cognitive dissonance. Similarities between cognitive dissonance and felt ambivalence have been noted by several researchers (e.g., Jonas, Broemer, & Diehl, 2000; McGregor, Newby-Clark, & Zanna, 1999) – both concepts describe aversive feelings caused by being aware of incongruence of one's beliefs regarding an attitude object. The main difference between these two concepts concerns the situations by which they are caused (van Harreveld, van der Pligt, & de Liver, 2009). While felt ambivalence arises in situations in which attention is directed at an ambivalent attitude, cognitive dissonance arises in situation in which a univalent attitude is disturbed (e.g., by inducing behavior incongruent with an individual's attitude; Festinger & Carlsmith, 1959). However, the consequences of felt ambivalence and cognitive dissonance are similar. This point is illustrated by the similarities of two experiments focused on the role of arousal in dissonance reduction (Zanna & Cooper, 1974) and on biased information processing serving ambivalence reduction (Study1; Nordgren et al., 2006), respectively. In both experiments, participants were first administered a sugar pill, but were either told that the pill would make them feel aroused or relaxed. The results in both experiments were similar – when participants were told that the pill would be relaxing they showed dissonance reduction and biased information processing. In contrast, when they were told that the pill was arousing, participants showed neither dissonance reduction nor biased information processing. Both Zanna and Cooper (1974) and Nordgren et al. (2006) argue that the reason for this pattern of results is that participants attributed their negative feelings caused by cognitive dissonance or ambivalence to the effects of the pill. We take the results of these experiments as indication that negative feelings caused by cognitive dissonance and ambivalence both in fact result from attitudinal entropy, with the difference being that in cognitive dissonance paradigms entropy is induced and in ambivalence paradigms attention to high entropy attitudes is induced.

Prediction 13: Given that the AE framework assumes that felt ambivalence and cognitive dissonance are caused by aversive configurations of the attitude network in combination with high dependence, felt ambivalence and cognitive dissonance are predicted to have similar consequences.

6.5 Future Study of the Attitudinal Entropy (AE) Framework

In the remainder of this chapter we want to address some important opportunities for future study of the AE framework. Apart from the empirical predictions that follow from the AE framework, we want to highlight the possibility of finding neural substrates of the AE framework's propositions, possibilities for further theoretical integration, and open questions raised by the AE framework.

6.5.1 Possible Neural Substrates of the AE Framework

Affective neuroscience has identified several neural substrates of attitude dynamics. Much of this research has focused on finding neural substrates of the reaction to valenced stimuli. This research has identified that the amygdala plays a central role in processing valenced stimuli (e.g., Morris, Frith, Perrett, & Rowland, 1996; Phelps, 2006; Zald, 2003). Importantly, the amygdala seems to integrate information from throughout the brain (Cunningham & Zelazo, 2007), which makes it likely that global evaluations are formed in this neural structure. Another neural structure that plays a central role in attitude dynamics seems to be the anterior cingulate cortex (ACC). The ACC plays an important role in the detection of potential conflict (Carter et al., 1998) and it was also shown that the ACC is active during the experience of cognitive dissonance (van Veen, Krug, Schooler, & Carter, 2009) and when ambivalent stimuli are processed (Cunnningham, Raye, & Johnson, 2004). This makes the ACC a likely candidate for the neural structure involved in translating entropy of attitudes under high dependence into aversive feelings (note that also other neural substrates are likely to be involved in the processing of ambivalent stimuli, such as the insula, the temporal parietal junction and the posterior cingulate cortex; see Nohlen, van Harreveld, Rotteveel, Lelieveld, & Crone, 2013).

Because the AE framework proposes that directing attention to and thinking about attitude objects serves the function of reducing attitudinal entropy, research on the neural substrates of consciousness is relevant to the AE framework. A recent influential theory of the neural underpinnings of consciousness posits that conscious experience results from neurons engaging in recurrent processing of stimuli, which enables information exchange between several low- and high-level areas of the brain (Block, 2005, 2007; Lamme, 2003, 2006). It thus seems likely that conscious processing of attitude objects results from integrating different kinds of information regarding the attitude object. This idea is also in line with the Information Integration Theory of consciousness (IIT; Tononi, 2004; Tononi & Edelman, 1998), which holds that the level of a system's consciousness depends on the amount of information this system integrates. This again underscores the importance of conscious thought in information integration. Information integration in turn is an important requirement for entropy reduction, thus further supporting the AE framework's assumption that a central function of conscious thought is to reduce attitudinal entropy.

6.5.2 The AE Framework's Relation to Other Models of Attitude

While it is beyond the scope of the current chapter to discuss the AE framework's relation to all prominent models of attitude, we discuss the AE framework's relation to three models that are in our view especially relevant: the Iterative Reprocessing (IR) model (Cunningham & Zelazo, 2007), the Attitude as Constraint Satisfaction (ACS) model (Monroe & Read, 2008) as an exemplar of constraint-satisfaction based connectionist models, and the Associative Propositional Evaluation (APE) model (Gawronski & Bodenhausen, 2006). These models are especially relevant, because they are similar in focus as the AE framework. The basic assumption of the APE model is that evaluations tapped by implicit measures result from associative processes, while evaluations tapped by explicit measures result from propositional processes. The APE model further assumes that cognitive consistency is only relevant to propositional processes. Similarly, the AE framework holds that attitudinal entropy reduction, which is mostly pronounced during explicit processing of the attitude object, results in heightened cognitive consistency. However, the models diverge in the assumption that heightened cognitive consistency during explicit processing of the attitude object results from a process that is qualitatively different from implicit processing of the attitude. In this sense, the AE framework is more in line with the IR model and the ACS model, which both assume that implicit and explicit evaluations are based on the same processes.

As we mention in the introduction of the AE framework, the process by which complex attitudinal representations are reduced to a single global evaluation is partly based on the IR model, which assumes that global evaluations are the result of iterative reprocessing of the attitude object, serving the reduction of entropy (Cunnningham, Dunfield, & Stillman, 2013). The AE framework shares also several similarities with the ACS model, as both models assume that the main driving factor in attitude dynamics is the drive for cognitive consistency. The ACS model and the AE framework also share a more technical similarity, because the ACS model is based on Hopfield (1982, 1984) neural networks, which in turn are based on Ising models. In our view the ACS model and the AE framework are therefore likely to complement each other and have different weaknesses and strengths. A strong feature of the ACS model is that it provides a formalized account of evaluative learning, while the AE framework is more parsimonious than the ACS model, which in our view has two advantages: First, parsimony aids the objective of ”understanding by building”, in the sense that the more parsimonious the model, the more likely it is that we can come to an understanding of the modeled construct. Second, parsimony also aids the development of predictions, because parsimony of a model makes it also less variable. Ultimately, we think that important knowledge can be gained by integrating these different models of attitudes. Based on the similarities between the IR model, the ACS model, and the AE framework, we are optimistic that such integration is possible (for an integration of the IR model and the ACS model see Ehret et al., 2015). As discussed in the introduction of the AE framework we are currently working on such integration.

6.5.3 Open Questions

The AE framework fosters subsequent research on attitudes in two different ways. First, as we discuss throughout this chapter, several predictions can be straightforwardly derived from the AE framework. Second, the AE framework also identifies several open questions, which we will discuss next.

Open Question 1: The exact nature of attitude elements needs to be further investigated. In our earlier work on attitude networks (Dalege et al., 2016; Dalege, Borsboom, van Harreveld, Waldorp, & van der Maas, 2017; Dalege, Borsboom, van Harreveld, & van der Maas, 2017; Dalege et al., 2018b) we treated rather general beliefs (e.g., judging a presidential candidate as honest) and feelings (feeling anger toward a presidential candidate), as well as concrete behaviors (voting for a presidential candidate) as attitude elements. However, it might also be possible that more low level beliefs (e.g., episodic memories of a person acting in a specific way) and feelings (e.g., recalling situations in which a person made one feel in a given way) are alternative operationalizations of attitude elements.

Open Question 2: While we have focused on determinants of entropy reduction, it is also relevant to investigate determinants that make individuals more tolerant to attitudinal entropy. A possible such determinant might be that individuals are highly motivated to be accurate.

Open Question 3: Can one level of attitudinal entropy reduction substitute for the other (e.g., is commitment to a given evaluation always necessary to reach higher levels of attitudinal entropy reduction or would, for example, relevance of the attitude to a decision be sufficient)?

Open Question 4: The AE framework assumes that attitudinal entropy is evaluated through two processes – the energy of a given attitudinal configuration and the instability of an attitude. However, the extent to which these processes are linked is a matter for future research.

Open Question 5: While we discussed attitudinal entropy reduction mostly as an intrapersonal process, it is certainly also possible that there are interpersonal effects on attitudinal entropy reduction. A question that needs to be addressed is how often individuals spontaneously reduce attitudinal entropy compared to how often this is socially instigated.

Open Question 6a: How pronounced are individual differences in attitudinal entropy reduction? Indirect evidence points to the existence of substantial differences, as individuals differ in their preference for consistency (Cialdini, Trost, & Newsom, 1995).

Open Question 6b: How pronounced are cultural differences in attitudinal entropy reduction? Similar to Open Question 6a, indirect evidence supports the hypothesis that cultural differences exist in how routinely individuals engage in attitudinal entropy reduction, as individuals from collectivistic cultures are less likely to experience cognitive dissonance than individuals from independent cultures (Heine & Lehman, 1997; Hoshino-Browne et al., 2005; Kitayama, Snibbe, Markus, & Suzuki, 2004).

Open Question 6c: Combining Open Questions 6a and 6b leads to the question whether individuals might even differ qualitatively in attitudinal entropy reduction: Are there individuals who do not engage in attitudinal entropy reduction?

Open Question 7: In the current chapter we focused on single attitudes. Attitudes, however, do not exist in independence from each other and future study of the AE framework should explore whether its principles also extend to interattitudinal processes.

6.6 Conclusion

In this chapter, we introduced the AE framework, which holds that (i) attitude inconsistency is entropy, (ii) energy of attitude configurations serves as a local processing strategy to reduce the global entropy of attitude networks, and (iii) directing attention to and thinking about attitude objects reduces attitudinal entropy by increasing the dependence parameter of attitude networks. The level of attitudinal entropy reduction depends on several factors, with merely directing attention to and thinking shortly about the attitude object representing the initial levels. Thinking more elaborately about an attitude object and commitment to an evaluation and relevance to decisions of the attitude represent the intermediate levels and high attitude importance represents the final level in attitudinal entropy reduction. We discussed the AE framework's relevance to research on ambivalence, the mere thought effect on attitude polarization, attitude strength, heuristic versus systematic persuasion, and implicit versus explicit measurements of attitude, thereby underscoring the integrative power of the AE framework. We also discussed several predictions that follow from the AE framework and several open questions identified by the AE framework. It is our view that because of its abilities in integration and spurring novel research questions, the AE framework represents a significant advancement in the theoretical understanding of attitudes. Furthermore, the AE framework places attitude dynamics into a broader dynamical systems context, further underscoring that reduction of entropy is the defining feature of living systems – both in a biological as well as a psychological sense. Ultimately, this might help to answer the question *why* it is that we think: to reduce the entropy of our mental representations.

Chapter 7

The Attitudinal Entropy (AE) Framework: Clarifications, Extensions, and Future Directions

Abstract

Does the Attitudinal Entropy (AE) framework provide a general theory of individual attitudes? This reply addresses several topics raised by the commentators. We first discuss the aim of the AE framework and provide some clarifications regarding its central principles. Then we evaluate how well the AE framework fares in explaining findings from dual-process approaches to attitudes and discuss the novelty of the AE framework's predictions. Finally, we apply the AE framework to psychological measurement and the interpersonal realm and provide an online app in which the basic workings of the AE framework are implemented.

This chapter has been adapted from: Dalege, J., Borsboom, D., van Harreveld, F., Lunansky, G., & van der Maas, H. L. J. (2018). The Attitudinal Entropy (AE) framework: Clarifications, extensions, and future directions. *Psychological Inquiry*, *29*, 218-228.

7.1 Prelude

Chapter 7 was written in response to several commentaries on Dalege et al., 2018a (Chapter 6). Though Chapter 7 mostly contains novel ideas and extensions to Dalege et al., 2018a, many of the ideas and extensions build on the topics raised by the commentators. To provide a comprehensive picture of the exchange between the commentators and us, I provide a summary of the main points of each commentary prior to Chapter 7.

7.2 Commentaries

Daley, M. (2018). Circe's victims: Are we too easily seduced by the siren song of mathematical physics? *Psychological Inquiry, 29,* 194-195.

Daley discusses the use of applying concepts from physics to psychology. He argues that physics is much simpler than psychology and cautions against too leniently applying physics concepts to psychology. In his view, applying the Ising model to attitudes might create the illusion of explanation, because the Ising model is capable of universal computation (i.e., any imaginable function can be implemented in the Ising model). Based on the fact that the Ising model is capable of universal computation, Daley argues that simulating attitude dynamics using the Ising model does not provide an argument for the correctness of the Attitudinal Entropy (AE) framework, which rests on the Ising model. However, further study of the AE framework's implementation in the Ising model might provide interesting hypotheses for future research.

March, D. S., Olson, M. A., & Gaertner, L. (2018). On logical and mathematical boxes: Does the attitudinal entropy framework expand our understanding of attitudes? *Psychological Inquiry, 29* 196-199.

March, Olson, and Gartner think that the Attitudinal Entropy (AE) framework provides an unorthodox approach to attitudes, which might challenge conventional approaches to attitudes. They, however, argue that to do so the AE framework needs to be better aligned with current research on attitudes and they question the conceptual novelty of the AE framework. They first discuss the definition of micro- and macrostates in the AE framework and argue that it is unclear whether attitude elements at the microlevel (i.e., beliefs, feeling, and behaviors vis-á-vis an attitude object) have influence on each other and whether cognitive consistency is increased at this level. Similarly, they wonder whether macrostates of attitudes (i.e., summaries of the attitude elements) take tension between positive and negative evaluations into account. Regarding entropy, March et al., wonder what the value is of defining attitude consistency as Boltzmann entropy and attitude stability as Gibbs entropy. They are also critical of using energy as an analogy for attitudinal processes and wonder whether the AE framework takes the difference between neutral and ambivalent attitudes into account and whether the AE framework can distinguish between different bases of attitudes.

March et al. then evaluate the AE framework's contribution to the conceptual understanding of attitudes. They first question the AE framework's assumption that measurement influences attitudinal properties and argue that the AE framework does not take differences between implicit and explicit measures, such as differential prediction of behavior and reliability, into account. They then argue that the AE framework treats attitudes as states rather than traits, which in their view is appropriate for weak attitudes but not for strong attitudes. On the other hand, the AE framework shows promise in describing change of strong attitudes, which according to March et al. is an overlooked issue in much of current research on attitudes. Regarding conceptual novelty, March et al. argue that the AE framework is similar to cognitive dissonance theory as the central assumptions of these theories align well. Furthermore, they also argue that some of the hypotheses derived from the AE framework were already tested or are trivial. March et al. conclude that it is more laudable for a theory to make novel predictions than to account for well-established findings and that the AE framework does fare well in the latter but not in the former.

Monroe, B. M. (2018). Progress in understanding requires modeling with faithfulness to the phenomenon, and precise mechanisms. *Psychological Inquiry, 29*, 200-202.

Monroe thinks that creative attempts, like the formulation of the Attitudinal Entropy (AE) framework, are a necessary way forward for the advancement of theorizing in social psychology. He, however, questions some of the AE framework's core assumptions and argues that his Attitudes as Constraint Satisfaction (ACS) model does a better job in modeling attitudinal processes. Monroe first argues that the AE framework does not take learning of associations between attitude elements (i.e., beliefs, feeling, and behaviors vis-á-vis an attitude object) and the attitude object into account and that because of this all attitude elements are equally important. In contrast, the ACS model takes learning into account based on the updating of association weights when the attitude object is processed and can therefore account for associations differing in strength. According to Monroe, because of this learning process, the ACS model captures the integration of implicit and explicit attitudes. In the ACS model, implicit attitudes represent the outcome of the first recognizable signal in the processing of the attitude object and explicit attitudes represent the outcome of further processing. In Monroe's view this process provides a more realistic explanation of the differences between implicit and explicit attitudes than the process that the AE framework assumes. The AE framework models implicit attitudes by assuming that implicit measurement of attitudes results in more randomness in the behavior of the attitude elements than explicit measurement of attitudes. According to Monroe this is not realistic, because it does not take the different time intervals of these measurements into account.

Monroe then contrasts how state updating is implemented in the ACS model with how it is implemented in the AE framework. While the ACS model uses the way neurons communicate with each other to model interactions between attitude elements, the AE framework uses Glauber dynamics to model these interactions. According to Monroe the causal process of how attitude elements influence each other is more clearly defined in the ACS model than in the AE framework. He also argues that the way the AE framework models attitude change is not satisfying, because changing thresholds in the Ising model is akin to changing attitude elements for no reason. Additionally, he criticizes that the variations in the AE framework's dependence parameter do not correspond to the phenomenology of thinking, because in his view it seems unreasonable that attitude elements change in a stochastic fashion. In contrast, the ACS model implements thinking about an attitude object by increases in the activation of the node representing the attitude object. Finally, Monroe argues that the justification for the way the AE framework models ambivalence is unclear and that simpler modeling of ambivalence would be possible. Monroe, however, sees value in the AE framework's assumption that attitude change can take place in a catastrophic fashion. He concludes that while the AE framework is more precise than most models in social psychology, it needs to provide more realistic implementations of attitude dynamics.

Petty, R. E., & Briñol (2018). The AEF: Reinforcing our knowledge about attitudes using a physics metaphor. *Psychological Inquiry, 29*, 203-207.

Petty and Briñol applaud that the Attitudinal Entropy (AE) framework integrates several well-established findings in the attitude literature, such as the positive relation between elaboration and attitude strength. They, however, also identify some issues with the AE framework. First, Petty and Briñol argue that the AE framework focuses too much on inconsistent and unstable attitudes, because the AE framework assumes that inconsistency and instability are natural states of attitudes. Second, they argue that the AE framework assumes that stability and consistency always co-occur, because they are both defined as attitudinal entropy. In contrast, it might be more beneficial to conceptualize inconsistency as an antecedent of attitude strength and stability as an outcome of attitude strength.

Petty and Briñol then discuss how the AE framework treats the processing of heuristic cues and argument quality under high and low involvement. They argue that the AE framework is able to model the general phenomenon that attitude change depends more on heuristic cues under low involvement, while attitude change under high involvement depends more on quality of arguments. A dual-process explanation of attitude change, however, is needed to account for more specific effects of persuasion and for the cognitive mechanisms leading to attitude change. Based on the Elaboration Likelihood Model (ELM), Petty and Briñol argue that heuristic cues, like source attractiveness, can lead to attitude change under low and high involvement, but the process by which they do that is different under low versus high involvement. These different processes then lead to weak attitudes under low involvement, while they lead to strong attitudes under high involvement. In addition, Petty and Briñol argue that the AE framework does not address the finding that under high involvement many weak arguments lead to less attitude change than few weak arguments.

Petty and Briñol also argue that the AE framework treats cognitive consistency in a too simplified way and that it does not address other processes than thinking that reduce inconsistency. Additionally, they address some areas of confusion. First, they argue that the AE framework's assumption that implicit measures tap inconsistent and unstable attitudes does not agree with some prominent models on implicit measurement, which assume that implicit measures tap stable and consistent attitudes. Second, while they judge the AE framework's prediction of an opposite mere thought effect (i.e., having little time to answer attitude questions leads to less polarized attitudes) as intriguing, they argue that it is unlikely to be correct. A rushed response might lead individuals to focus only on one side of an ambivalent attitude, which would lead to higher polarization. Petty and Briñol conclude that the AE framework offers a high degree in precision and elegance, but is less nuanced than other theories on attitude. While sometimes it is not clear whether some of the predictions of the AE framework could not be derived from other theories, the AE framework does provide novel predictions and ideas and therefore might advance the attitude literature.

Rios, K., & Roth, Z. C. (2018). Extending the Attitudinal Entropy Framework to Interpersonal Phenomena: A Commentary on Dalege et al. (2018). *Psychological Inquiry, 29,* 208-212.

Rios and Roth apply the principles of the Attitudinal Entropy (AE) framework to interpersonal phenomena. They first discuss how interpersonal processes affect attitudinal entropy. Generally, attitudinally heterogeneous social networks reduce certainty and increase ambivalence of individuals belonging to these networks, resulting in higher attitudinal entropy. Similarly, individuals holding a minority opinion report lower certainty and higher ambivalence. Sometimes, however, individuals holding a minority position hold attitudes high in certainty and low in ambivalence and then are also willing to express their opinion. Rios and Roth then link conformity to entropy reduction and argue that group processes regarding entropy reduction are similar to entropy reduction at the individual level. For example, when a group directs attention at a specific issue conformity on this issue is likely to increase.

Rios and Roth then turn to interpersonal determinants of the motivation to reduce attitudinal entropy. Holding consistent and stable attitudes helps in having smooth social interactions and protects one from being seen as a hypocrite. Similarly, groups in which members disagree with each other (i.e., high entropy at the group level) lead to discomfort and ambivalence and individuals generally strive to reduce entropy at the group level. There are, however, also situations that make it likely that group members do not engage in entropy reduction at the group level. First, minority opinions can be functional, because individuals can express uniqueness through a minority opinion and holding a minority opinion high in certainty can support a clear self-concept. Second, individuals might be motivated to hold multiple social identities, which would represent high attitudinal entropy. Holding multiple social identities can help when one is motivated to identify with groups that hold different values. Third, in some instances individuals feel comfortable with holding conflicting attitudes when they are able to integrate seemingly conflicting attitudes. For example, religious scientists often express the opinion that faith and science are compatible and can even benefit each other. Finally, individuals might be motivated to shift their attitudes according to the social context. These shifts do not need to indicate randomness (and therefore high entropy) but can be strategic choices.

Rios and Roth also discuss the extent to which social context determines to which attitude objects individuals direct attention and thought. Individuals might be motivated to direct attention and thought at attitude objects to compare their own attitude with that of others to analyze whether they hold a 'correct' attitude. Accordingly, attitude certainty increases when individuals hold attitudes similar to the attitude of the majority or to the attitude of a high-status outgroup.

Apart from applying the AE framework's principles to interpersonal processes, Rios and Roth also argue that high entropy attitudes can have positive social consequences and that low entropy attitudes can have also negative consequences. Individuals, who hold attitudes with high certainty, tend to be unwilling to compromise when they engage with individuals, who hold opposite attitudes. In contrast, individuals who are uncertain have smoother interactions with individuals, who have opposite attitudes. Similarly, low entropy at the group level might be indicative of groupthink, which leads group members to only share information consistent with the opinion of the majority. Additionally, individuals who are able to shift their attitudes according to the social context are generally well liked.

Rios and Roth conclude that the AE framework can also be applied to interpersonal phenomena. By doing so, one can identify boundary conditions of the framework and one might be able to integrate intra- and interpersonal antecedents of attitude strength.

Van Dessel, P., De Houwer, J., Hughes, S., & Hussey, I. (2018). An analysis of the scientific status and limitations of the attitudinal entropy framework and an initial test of some of its empirical predictions. *Psychological Inquiry, 29*, 213-217.

Van Dessel, De Houwer, Hughes, and Hussey think that the Attitudinal Entropy (AE) offers a novel perspective on social psychology and that it provides a timely contribution, because of its focus on the importance of entropy in human cognition. They first evaluate to what extent the AE framework provides explanations or descriptions of attitudinal dynamics and argue that the AE framework is not situated at the cognitive level of explanation, but at the functional level of explanation and at the descriptive level. In their view, entropy is a descriptive concept and entropy reduction only concerns functional explanations. Similarly, they argue that the Causal Attitude Network (CAN) model, which is the predecessor of the AE framework, can be entirely interpreted in a functional fashion. The attitude elements making up attitude networks can all be conceptualized as (verbal) behavioral responses and therefore no references to mental processes have to be made. The CAN model and the AE framework can contribute to research on attitudes by illuminating how one behavior leads to other behaviors or how the environment determines behavior. Van Dessel et al. further argue that it might be possible to apply the AE framework to the cognitive level of explanation, but that this creates the danger of conflating behavior with mental processes. In their view, the AE framework can also be complemented by theories situated at the cognitive level to, for example, illuminate the motivation for attitudinal entropy reduction.

Van Dessel et al. then argue that the AE framework's specifications how attitude elements influence each other are limited and that the AE framework treats cognitive concepts, like motivation and inferences, too superficially. In their view, the AE framework does not provide an explanation for why individuals should be motivated to reduce entropy. Extending the AE framework to the cognitive level of explanation requires a more detailed analysis of inferential processes. In Van Dessel et al.'s inferential model of evaluative stimulus-action effects, inferences are based on action outcomes and provide a basis for the individual's mental model of the world. The AE framework and Van Dessel et al.'s model are largely compatible with each other. Both assume that entropy reduction is a crucial factor in attitude dynamics and both accounts do away with the idea of dual-processes and rely on a single process. In contrast to the AE framework, however, the inferential model makes a clear distinction between the functional and cognitive level of explanation by modeling evaluations and not attitudes. Additionally, the inferential model provides a more nuanced description of how inferences arise and how they influence evaluations and integrating the inferential model with the AE framework might lead to a better understanding of the cognitive processes involved in attitudinal entropy reduction.

Van Dessel et al. also tested two predictions of the AE framework. First, they tested whether internal consistency of implicit measures increases for attitude objects, individuals think frequently about. By analyzing scores on 190 different types of Implicit Association Tests (IAT), they found support for the AE framework's prediction. IATs assessing topics, individuals think frequently about, showed higher internal consistency than IATs assessing topics, individuals do not think often about. Second, they tested the prediction that, when individuals have to quickly answer attitude questions, extremity of attitudes decreases. They compared scores on attitude questions, for which individuals were instructed to provide their deliberative evaluations, to scores on attitude questions, for which individuals were instructed to provide their 'gut' evaluations. Gut evaluations were slightly more extreme than deliberative evaluations, which is not in line with the AE framework's prediction. Van Dessel et al. conclude that the support for one of the AE framework's prediction shows the predictive value of the AE framework and that the AE framework might provide a major advance in the attitude literature.

7.3 Introduction

In our target article, we formulated a general theory of attitudes – the Attitudinal Entropy (AE) framework (Dalege, Borsboom, van Harreveld, & van der Maas, 2018a). The AE framework rests on analogical modeling of attitudes using concepts from statistical mechanics and is built on three basic principles. First, attitude inconsistency and instability are two distinct but related indicators of attitudinal entropy – a measure of randomness and unpredictability derived from thermodynamics. Second, energy (in a metaphorical sense) of attitudinal configurations is used to locally evaluate the global entropy of an attitude. Third, attention and thought have an analogous effect on attitudinal entropy as (inverse) temperature has on thermodynamic systems – attention and thought reduce attitudinal entropy. In our target article, we show how this framework can be used to explain a host of established phenomena and to make novel predictions to test the theory empirically.

The commentaries on our target article generally agreed that the AE framework has potential to achieve its main task – capturing the immense complexity of attitude by a limited number of basic principles. This encourages us to further refine the AE framework, so that it will eventually become what we envision it can be: A formalized theory on attitudes that truly explains many established phenomena and that makes unambiguous predictions.

In this reply we take a first step in the refinement of the AE framework by addressing the commentaries' critiques. First, we discuss the value of formulating a formalized and parsimonious theory of attitudes and by this make the aim of the AE framework more explicit. Second, we take the opportunity to clarify the principles of the AE framework. Third, we show that the AE framework explains several findings from the dual-process literature in a more parsimonious way. Fourth, we discuss critiques regarding the question whether the AE framework makes truly novel predictions. Fifth, we discuss the implications for measurement of the AE framework, which hold that measurement fundamentally influences attitudinal processes. Sixth, we discuss some potential extensions of the AE framework to the interpersonal realm. Seventh, to make the formalisms of the AE framework more accessible and applicable, we present an online app in which the basics of the AE framework are implemented.

7.4 The Aim of the Attitudinal Entropy (AE) Framework

When developing the AE framework our aim was as simple as it was ambitious: the formulation of a theory of attitudes relying on few basic principles yet able to explain several of the complex findings in the vast attitude literature. In our view only such a theory is a practical one: The more parsimonious and constrained a theory is, the more concrete predictions it makes, the more understanding of the modeled system it provides and the more promise for eventual control over the modeled system it holds. While we do not think that the AE framework is the only possibility to arrive at such a theory of attitudes, we do hold that psychology would do well to invest more energy in the development of formalized theories that can be used to explain, model, simulate, and analyze empirical phenomena.

One crucial test for any general theory is whether it can explain established phenomena. A central aim of our target article was therefore to provide a first intensive investigation of whether the AE framework is up to this task. Some commentators were not entirely convinced by the value of this aim. For example, while March, Olson, and Gaertner (2018) are supportive of our aim to derive a formalized theory on attitudes, they state that "[i]t is far more valuable... when such models make novel predictions [than to explain established findings]". However, in our view investigating whether a theory is able to explain established phenomena is important, because, any theory that *only* makes novel predictions without accounting for established findings is in danger to contribute to a highly fragmented research field, because it does not specify how these novel predictions relate to established phenomena (e.g., Kruglanski, 2001; Vallacher & Nowak, 1997).

Clarifying the aim of the AE framework seemed also important to us, because unfortunately Daley (2018) misunderstood the aim of the AE framework. While he sees value in linking attitude research to the literature on entropy, he argues that implementing the AE framework in the Ising (1925) model is not a compelling argument for the correctness of the AE framework. Of course, it is true that implementing the AE framework in the Ising model by itself does not support the correctness of the AE framework, as Daley correctly suggests. However, we never made such a claim. Instead, we implemented the AE framework in the Ising model, because this model lends itself very well to the principles of the AE framework and therefore represents a straightforward opportunity to formalize these principles. Thus, the Ising model implements basic principles, as articulated in the AE framework, which can be motivated on independent substantive terms.

The fact that the Ising model can be used to implement these principles in such a straightforward fashion is, in our view, surprising and noteworthy. However, of course we do not interpret the fact that the AE framework can be implemented in this way as evidence for the model's empirical adequacy, as Daley appears to suggest. Support for the correctness of the AE framework instead comes from the fact that it can successfully reproduce established phenomena. Another point of confusion apparent in Daley's commentary concerns the universality of the Ising model. While the universality of the Ising model poses a problem for underspecified theories, the AE framework considerably constrains the behavior of the Ising model by providing a clear mapping between the mathematical parameters in the Ising model and the substantive interpretations of these parameters. For example, the main principle of the AE framework that attention and thought reduce attitudinal entropy considerably reduces the flexibility of the Ising model, because low attention *must* result in more random and unstable behavior by keeping the dependence parameter (i.e., the parameter used to model attention and thought) in the Ising model at a low value. Thus, while the unconstrained Ising model can definitely fit not just one elephant, but an entire herd of them, as Daley correctly suggest, this does *not* mean that the constrained Ising model we use can fit any conceivable data. On the contrary, the model has clear truth conditions and falsifying instances; for example, if the mere thought effect turned out not to exist (under the conditions specified by the AE framework), the AE framework would be in serious trouble.

An interesting issue regarding the level of explanation that the AE framework provides was raised by Van Dessel, De Houwer, Hughes, and Hussey (2018). They argue that the AE framework has much value at the descriptive level and at the functional level of explanation, but that its cognitive explanations are limited. Indeed, the focus of the AE framework lies at the level of functional explanations and we agree with Van Dessel et al. that future study of the AE framework can focus on providing less abstract cognitive explanations of its central principles (e.g., what is the cognitive process by which attention and thought reduce attitudinal entropy?). In our view, however, Van Dessel et al. somewhat undervalue the cognitive explanations the AE framework already provides. They, for example, argue that the AE framework does not provide an explanation for the motivation to reduce entropy other than that it causes distress. While it seems that we did not communicate this issue optimally in our target article, there are ample reasons why individuals are motivated to reduce attitudinal entropy. For example, the functions typically associated with attitudes cannot be fulfilled by high-entropy attitudes. Based on this argument, it is our view that the AE framework fairs fairly well in explaining the motivation to reduce attitudinal entropy. A related question, however, might be addressed by a more detailed cognitive explanation: What determines that individuals direct attention and thought at some attitudes (and thereby lower the entropy of these attitudes) but not others?

7.5 Principles and Definitions of the AE Framework

The central principle of the AE framework holds that inconsistency and instability reflect attitudinal entropy, which results in attitudes naturally being driven towards inconsistency and instability. While Petty and Briñol (2018) agree that this principle generates an interesting connection between psychological processes and physical laws, they worry that this principle leads us to focus too much on instable and inconsistent attitudes. We are glad that Petty and Briñol pointed out this issue, because we did not want to suggest that the AE framework focuses on instable and inconsistent attitudes, but that linking inconsistency and instability of attitudes to entropy has implications not only for inconsistent and unstable attitudes but also for consistent and stable attitudes – and all other attitudes, for that matter. The crucial point of the AE framework's central principle is that humans are in a constant struggle of keeping their attitudes from drifting to their natural state of high entropy. The AE framework therefore can account for unstable and inconsistent attitudes (high entropy attitudes to which the individual does not direct much attention and thought), very stable and consistent attitudes (low entropy attitudes that are constantly on the individual's mind) and attitudes anywhere in between (moderate entropy attitudes to which individuals direct some attention and thought). In fact, according to the AE framework, all of these variants of attitudes *must* exist, because a given individual cannot direct attention and thought to all attitude objects at the same time.

Petty and Briñol (2018) further question whether defining stability and inconsistency of attitudes as related indicators of attitudinal entropy leads to the theoretical lumping of stability and consistency of attitudes. We are glad for being able to clarify this point, because we agree that lumping stability and consistency would not be a good idea. While the AE framework does indeed hold that inconsistency and instability are both indicators of entropy, this does not imply that inconsistency and instability are the same thing; it merely means that they are related constructs. As we stated in our target article, low Gibbs entropy (which relates to stability of attitudes) *creates* the possibility for low Boltzmann entropy (which relates to consistency of attitudes). Thus, from the perspective of the AE framework, stability can exist independently from consistency (attitudes can be stable and ambivalent), but consistency cannot exist without stability.

Another point that requires some clarification is what micro- and macrostates specifically represent in the AE framework. March et al. (2018) argue that defining microstates of attitudes as the elements making up the attitude would better align with classic expectancy-value models (e.g., Fishbein & Ajzen, 1975) than our definition. We apparently did not communicate this issue optimally, as our definition aimed to do exactly that: Definition I of the AE framework holds that "The configuration of the attitude elements constitutes the microstate of the attitude" where, as we also explicitly stated, the configuration of an attitude depends on the specific states of the given elements. We intentionally used this definition *because* it aligns well with classic theories on attitude structure, like expectancy-value models. Our definition of attitudinal microstates is thus exactly in accordance with what March et al. advocate.

Similarly, March et al. (2018) provide some arguments that seemingly go against the second definition of the AE framework, which holds that the macrostate of an attitude is represented by the number of positive vs. negative attitude elements. They argue that this definition of macrostates does not take relationships between elements into account. We are afraid that, again, we may not have communicated the theory optimally, because in a nontrivial sense, the relations between attitude elements in a network structure are the explanatory workhorse of the theory (see also Dalege et al., 2016) – in fact, in the AE framework, increasing the dependency between attitude elements is the *main* process that reduces attitudinal entropy. March et al. also claim that the AE framework does not capture the potential tension between negative and positive elements at the macrolevel, which is contrary to the formulation of the theory: The tension between negative and positive elements is actually represented by the entropy of the attitudinal macrostate – the higher the number of conflicting attitude elements, the higher the Boltzmann entropy.

Another interesting question raised by some commentators was whether the AE framework could integrate different bases of attitudes (March et al., 2018) and differing importance of attitude elements (Monroe, 2018). While the simplified Ising models we use in our target article do not take these issues into account – in fact, each node in these models is exchangeable – the Causal Attitude Network (CAN) model (Dalege et al., 2016), on which the AE framework is based, does integrate these issues. First, structural importance of nodes (i.e., centrality) is expected to reveal different bases of attitudes (e.g., affective versus cognitive bases; c.f. Edwards, 1990; Fabrigar & Petty, 1999). For example, in the CAN model a network in which affective nodes have more connections (and are therefore more central) than cognitive nodes represents an affect-based attitude. We refer the interested reader to Dalege et al. (2016) and Dalege, Borsboom, van Harreveld, and van der Maas (2017) for an in-depth discussion of the meaning of centrality in attitude networks. Second, Monroe is correct in pointing out that the importance of attitude elements is not directly weighted in the contribution to the attitude's macrostate within the AE framework. While it might eventually become apparent that the AE framework needs to explicitly take weighting of attitude elements into account, structural importance of attitude elements already indirectly leads to differences in the contribution to the attitude's macrostate. As a simple example, take a network with four nodes – nodes 2 and 3 are both strongly connected to node 1 and node 4 is not connected to any node. Node 1 is thus more structurally important than node 4 and the macrostate of the attitude is also more dependent on node 1 than node 4. The reason for this is that the states of nodes 2 and 3 are largely dependent on the state of node 1, so if node 1 is negative (positive) it is likely that nodes 2 and 3 are also negative (positive). The state of node 4, on the other hand, has no impact on the states of other nodes and is therefore less predictive of the macrostate of the attitude.

Another point that requires additional clarification is the process through which a given node changes its state due to the influence of other nodes. In his commentary, Monroe (2018) argues that whether a node changes boils down to "a probabilistic roll of the dice". This statement, however, does not fully encompass our proposed process through which nodes change (or remain in) their state. The updating of states in the AE framework is based on Glauber (1963) dynamics and the basic workings of this process are the following. During each iteration a node is randomly chosen. Then the current energy of this node is calculated. This energy depends on the extent to which the node's state aligns with the connection the node has to other nodes and with its threshold (the disposition of the node). Then the energy of the node is calculated for its opposite state. If the node's opposite state is lower in energy than its current state, the node is likely to change its state and the likelihood depends on the difference in energy and on the value of the dependence parameter. The higher the dependence parameter, the more deterministic the system becomes. Only under a low dependence parameter is the changing of a node's state comparable to a roll of dice. In our view this is one of the strengths of the AE framework: It can account for random and unpredictable behavior *but* also for organized and predictable behavior.

In general, while Monroe (2018) judges the attempt of the AE framework to explain attitudinal phenomena in a more formalized and rigorous way as a necessary step towards theoretical progress, he also questions how realistic some modeling choices of the AE framework are. He contrasts the AE framework with his connectionist neural network model of attitudes – the Attitudes as Constraint Satisfaction (ACS) model (Monroe & Read, 2008) – and argues that the ACS model is a more realistic model of attitude dynamics than the AE framework. We agree that the ACS model fairs well in providing a realistic model of how dynamics of neural networks relate to attitude dynamics. However, a realistic model of any given process is not necessarily a good theory. For example, neural network models hardly fulfill the central aim of theorizing to make the basic elements of the theory as simple as possible. This does, of course, not mean that they are not useful – to the contrary, building a copy of the brain can deliver many insights, but to truly build a theory on these insights one needs to simplify the modeled processes. The AE framework represents such an attempt and is therefore also complementary rather than in competition with neural network models like the ACS model. One can think of the AE framework as the attempt to provide a map of relevant attitudinal processes, while neural network models try to rebuild the whole planet of attitude processes. In our view the additional value of neural network models is that they can be used to investigate where more parsimonious theories like the AE framework break down. A fruitful avenue for integration of the AE framework and neural network models like the ACS model would therefore be a detailed investigation which predictions are provided by the ACS model that do not follow from the AE framework. One area where the ACS model probably has more predictive power than the AE framework is the learning of attitudes. Currently, the AE framework does not specify how attitude networks develop and ongoing investigations of our lab focus on integrating Hebbian learning as used in the ACS model into the AE framework.

7.6 Beyond Dual-Process Models

A common theme in several commentaries was the question whether the AE framework is able to integrate insights from dual-process models. While Van Dessel et al. (2018) were in support of the AE framework as an alternative to dual-process models, Petty and Briñol (2018) and March et al. (2018) were not entirely convinced about the explanatory scope of the AE framework regarding findings from the dual-process paradigm. In our view the AE framework does a good job in explaining findings from the dual-process tradition. In the target article we already modeled a central finding from the dual-process approach to persuasion and in this section we first further explore how well the AE framework fares in explaining findings from this literature. We then show that the AE framework is also able to explain state-like and trait-like behavior of attitudes.

We were very glad that two prominent researchers from the dual-process tradition in persuasion research agreed that the AE framework is able to explain the general finding that low involvement leads to heuristic processing while high involvement leads to systematic processing (Petty & Briñol, 2018). Petty and Briñol, however, also challenge us to explain more nuanced effects from the dual-process tradition and we are glad for the opportunity to do so. Specifically, they argue that the AE framework does not predict the finding that under low involvement number of arguments determines attitude change, while under high involvement quality of arguments determines attitude change (Petty & Cacioppo, 1984). Because the AE framework is a formalized theory we are able to test whether this claim is correct.

7.6.1 Simulation R1: Modeling Impact of Argument Quantity Versus Argument Quality

To test whether the AE framework predicts the effects of quantity versus quality of arguments under low versus high involvement, we set up a simulation similar to Simulation 3 of our target article. As in Simulation 3, we first created a low involvement group represented by a low dependence parameter ($\beta = 1$) and a high involvement group represented by a high dependence parameter ($\beta = 3$) and initialized all individuals' attitudes to have a positive disposition (all thresholds = 0.2). We straightforwardly modeled the effect of number of arguments by changing either the thresholds of three or nine nodes. The quality of arguments was modeled by different impact on the thresholds (weak arguments changed the thresholds to -0.2 and strong arguments changed the thresholds to -0.7). We then simulated 100 individuals in each of the six conditions (each individual was modeled using 1000 iterations of Glauber dynamics; in the first 500 iterations thresholds were set to the positive disposition to initialize the positive pre-persuasion attitude and in the second 500 iterations thresholds were set to the specific values of the given condition).

Figure 7.1 shows that the AE framework reproduced the global effect identified by Petty and Cacioppo (1984) and the three-way interaction on the sum score of the attitude elements at the $1,000^{th}$ iteration was significant, F (1, 792) = 7.52, p = .006, η_p^2 = .01. Figure 7.1a shows the influence of the different threshold change conditions in the low dependence parameter condition. In this condition a strong main effect is observed for the number of changed thresholds (reproducing the finding that under low involvement argument quantity mostly determines attitude change), F (1, 396) = 237.27, p < .001, η_p^2 = .37. While also significant, the main effect of strength of the threshold change was substantially lower (reproducing the finding that under low involvement argument quality has less impact on attitude change), F (1, 396) = 69.63, p < .001, η_p^2 = .15. We also observed a small interaction effect between number and strength of threshold change, F (1, 396) = 9.71, p = .002, η_p^2 = .02, which was mostly driven by the stronger impact of strength of threshold change when number of arguments was high. Figure 7.1b shows the influence of the different threshold change conditions in the high dependence parameter condition. In this condition a strong main effect is observed for the strength of the threshold change (reproducing the finding that under high involvement argument quality mostly determines attitude change), F (1, 396) = 429.93, p < .001, η_p^2 = .52. While also significant, the main effect of the number of changed thresholds was substantially lower (reproducing the

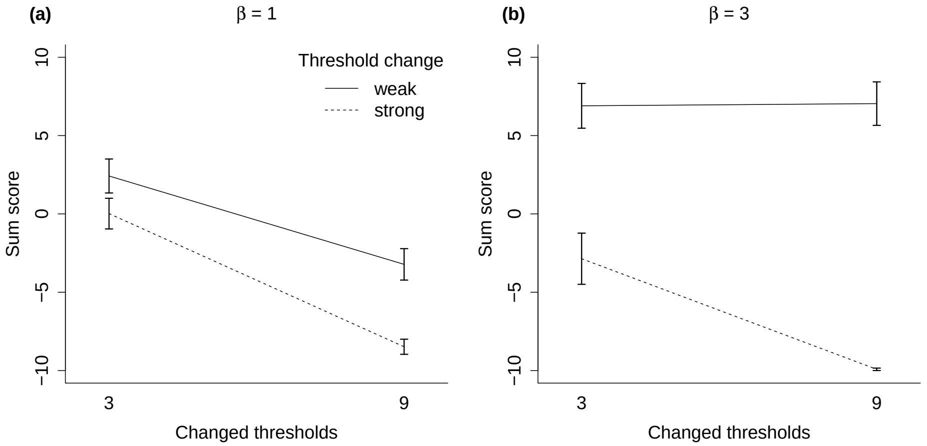


Figure 7.1: Effects of number and strength of threshold change under low (a) and high (b) dependence parameter. Error bars represent +- 2 standard errors around the mean.

finding that under high involvement argument quantity has less impact on attitude change), $F\ (1, 396) = 28.84$, $p < .001$, $\eta_p^2 = .07$. We also observed a small interaction effect between number and strength of threshold change, $F\ (1, 396) = 31.22$, $p < .001$, $\eta_p^2 = .07$, which was mostly driven by the stronger impact of strength of threshold change when number of arguments was high. The different effects of quantity versus quality of arguments under low versus high involvement thus in fact follow from the AE framework.

A noteworthy observation in addition to the global pattern of the results in Figure 7.1 is that the AE framework predicts that when weak arguments are administered to highly involved individuals, the number of arguments does not matter. If the arguments were too weak, persuasion would just not be effective no matter the number of arguments. In contrast, Petty and Briñol (2018) argue that under high involvement a large number of weak arguments leads to even *less* persuasion than a low number of weak arguments. This is thus indeed not what the AE framework would predict and a convincing empirical demonstration of this effect would indicate a limitation of the AE framework. A careful examination of the study by Petty and Cacioppo (1984) on which Petty and Briñol build their argument, however, reveals that the effect is far from robust – in fact, the effect did not even reach the conventional threshold for statistical significance so additional research is necessary to evaluate whether the effect in fact exists at all. Absent such evidence, in our view Simulation R1 lends further support to the claim that the AE framework can explain findings from the persuasion literature in a more parsimonious and constrained way than dual-process models like the Elaboration Likelihood Model (ELM; Petty & Cacioppo, 1986) and the Heuristic-Systematic Model (HSM; Chaiken et al., 1989).

March et al. (2018) use a dual-process approach to what they see as state-like and trait-like attitudes. In their view, weaker attitudes are more state-like, because they are fluctuating and highly context-dependent, while stronger attitudes are more trait-like, because they are stable and relatively immune to context effects. It is our view that such a description of attitudes has a lot of merit and that such characterizations represent the endpoints on a continuum. In fact, the AE framework suggests that attitudes range from being high in entropy (and therefore highly fluctuating and unpredictable) to attitudes being low in entropy (and therefore stable and predictable). Therefore, while March et al. are correct in stating that the AE framework treats attitudes as states, the AE framework is able to integrate theories assuming attitudes to be temporary constructions (e.g., Schwarz, 2007) with theories assuming attitudes to be stable representations in memory (e.g., Fazio, 2007).

From the perspective of the AE framework, whether attitudes are more state-like or more trait-like is determined by the dependence parameter. While the AE framework does away with the idea of a stable representation of a summary evaluation in memory, attitudes in low entropy states (i.e., attitudes to which individuals direct much attention and thought) behave exactly as what would be expected from a stored representation of the summary evaluation. Under a high dependence parameter, attitude elements keep each other in check and therefore the macrostate of the attitude remains in the same state. Thus, while in principle the macrostate of the attitude is consistently constructed, the construction process under a high dependence parameter always results in a similar global evaluation and would therefore *appear* as a stored representation. In our target article, we show that variations in the dependence parameter also explain variations in attitude strength, such as resistance to change and impact on behavior (see also Dalege et al., 2016, 2018b). Given that the AE framework provides a parsimonious explanation of both weak and strong attitudes, a dual-process explanation of strong and weak attitudes may be superfluous.

In conclusion, while dual-process models have considerably advanced the attitude literature by illuminating, for example, the different effects of persuasion under low versus high involvement, the value of the dual-process description may ultimately be heuristic rather than explanatory.

While indeed persuasion has different effects under low and high involvement, and while weak attitudes are quite different from strong attitudes, these differences may not arise because of different processes but rather reflect one and the same process operating under different conditions.

7.7 Does the AE Framework Provide Novel Predictions?

The commentators substantially diverged on the question whether the AE framework leads to novel predictions. On the one hand, Van Dessel et al. (2018) went so far as to actually test two predictions of the AE framework, Rios and Roth (2018) argue that the AE framework cannot only illuminate intrapersonal but also interpersonal processes and Petty and Briñol (2018) see the novel predictions as one of the AE framework's strengths. On the other hand, March et al. (2018) seem to argue the extremely strong thesis that the AE framework does not make any novel predictions at all, as they contrast the AE framework with "models [that] make novel predictions about heretofore unobserved phenomena". However, it is unclear what could lead March et al. to this conclusion; the only evidence appears to be that, in their view, two out of 17 predictions stated in our target article were either already tested or trivial; even if this is true, there would seem to be 15 predictions remaining. In this section, we present a detailed analysis on the question whether the AE framework makes novel predictions.

A first point regarding the novelty of predictions is that judging whether a prediction is novel seems to be a rather difficult task. As case in point, compare how Petty and Briñol (2018) judge the novelty of the AE framework's prediction that an opposite mere-thought effect exists (i.e., forcing individuals to quickly answer questions is expected to result in less extreme attitudes than when the questions are administered in an usual way) to March et al.'s (2018) judgment. While Petty and Briñol called this prediction "intriguing" and judged it to be counterintuitive and unlikely to be confirmed, March et al. called the prediction neither "new [n]or surprising" and are convinced that this prediction could easily be demonstrated.

Why do established experts in the attitude field diverge so strongly on whether or not an effect would likely be demonstrated? We suggest that one important reason that researchers working in the same area can differ so substantially in their judgment of whether a prediction is novel might be that evaluating predictions based on verbal theories results in far too much flexibility. By turning the meaning of words in their intended direction, both March et al. (2018) and Petty and Briñol (2018) are able to provide valid reasons for *both* the existence *and* non-existence of the opposite mere-thought effect. This, in our view, illustrates the Achilles' heel of current mainstream theorizing in psychology: its reliance on verbal stories makes theoretical predictions far too dependent on unarticulated interpretations of theoretical constructs. In our view, a crucial strength of the AE framework is that it provides predictions that *must* follow, and this also gives more value to these predictions, because the predictions follow *independently* of what the developers of the model might think. Such unambiguous predictions are in our view valuable, because testing such predictions represents a more diagnostic test than testing a prediction that is only loosely based on a theory and dependent on this or that theorist's interpretation of the model.

Interestingly, Van Dessel et al. (2018) conducted a first test of the opposite mere thought effect. They compared the extremity of attitude ratings, which asked individuals to report their reflected feelings, to the extremity of attitudes ratings, which asked individuals to report their gut feelings. Van Dessel et al. found that these gut feelings were slightly more extreme than reflected feelings (note, however, that this was a very small effect), which is not in line with the predicted opposite mere thought effect. However, the test by Van Dessel et al. relies on the strong (and in our view not always realistic) assumption that individuals have reliable knowledge about their high entropy attitudes. Given the assumption that the state of high attitudinal entropy arises when individuals do not pay attention to the attitude object, it would not seem evident that people have introspective knowledge about such attitude states. Thus, while the test of Van Dessel does provide indirect evidence against the opposite mere thought effect, we think that the jury is still out on whether this effect exists, and further, more direct empirical tests, are warranted.

Another issue that came up in the commentaries regarding the novelty of the AE framework's predictions was that some of them have already been tested and confirmed (March et al., 2018; Petty & Briñol, 2018). We acknowledge that a subset of the AE framework's predictions indeed have already been tested, and we thank the commentators for pointing this out. However, it is important to note that whether a prediction was tested before or after we derived the AE framework has no bearing on the predictive power of the AE framework or the degree to which the reported effects confer evidence on the theory: if it turns out that the Ancient Greeks already observed that heavy and light objects fall at the same speed, that should not diminish the evidential force of this finding vis-á-vis Galileo's theory of free fall. Thus, in our view, while the fact that some predictions of the AE framework have already been tested does show our unawareness of this part of the literature, this does not detract from the fact that such confirmations provide evidence for rather than against the AE framework.

To conclude, while the novelty of some of the AE framework's predictions might be debatable, most of the AE framework's predictions have potential to advance the attitude literature by (1) providing more precision than what could be derived from other theories (e.g., predictions regarding boundary conditions of the mere thought effect and the effectiveness of persuasion) and (2) by predicting entirely novel effects. Many of these effects are in the context of the measurement of attitudes and we discuss the AE framework's implications for measurement in detail in the next section.

7.8 Let's go All Psychology on Attitude Measurement

The probably most controversial implication of the AE framework is that measurement has fundamental influence on attitudes – the way attitudes are measured influences their entropy. While this implication met with doubts from some commentators (March et al., 2018; Petty & Briñol, 2018), other commentators appeared more enthusiastic about this implication and provided some first empirical support for the measurement implications of the AE framework (Van Dessel et al., 2018).

In contrast to March et al. (2018), who stated that their "knee-jerk reaction to [the measurement implications of the AE framework]... is that unless we've gone 'full-physics', where attitudes are akin to quantum particles whose very measurement affects them (as in Heisenberg's uncertainty principle), it risks false-reification to claim that measurement determines some attitudinal entropy", we think that one should keep an open mind about the possibility that measurement affects attitudinal processes unless convincing evidence accumulates for attitudes being immune to measurement effects. In our view, however, the evidence *for* measurement effects on attitudes is substantial. Measurement of attitudes, just as most psychological measurements, sets in motion a variety of processes (e.g., Cannell, Miller, & Oksenberg, 1981; Strack & Martin, 1987; Thurstone, 1927; Tourangeau, Rips, & Rasinski, 2000) and it is unlikely that this has no impact at all on the measured system. Furthermore, the reference to quantum physics by March et al. actually supports our theorizing, because there is strong empirical evidence that at least some cognitive processes follow quantum probabilities (e.g., Pothos & Busemeyer, 2013) and, especially relevant to the current discussion this also holds for measurement effects on judgments (e.g., Busemeyer & Wang, 2018; Wang & Busemeyer, 2013; Wang, Solloway, Shiffrin, & Busemeyer, 2014). Additionally, one of the pioneers of quantum mechanics – Niels Bohr – probably used ideas from William James on thought-processes to derive quantum mechanical principles (Stapp, 1993; Plotnitsky, 2012). The reason why the theory of quantum mechanics might be built on insights from William James on thought-processes is that the nature of the human mind lends itself well to the issues discovered in quantum physics. While in physics it is highly counterintuitive that particles are fuzzy and do not have a definite position in space, that they can interfere with themselves, or that measurement affects the nature of particles, it is immediately apparent by introspection that such concepts are very fitting for mental processes. Thoughts are fuzzy and often hard to pin down, thinking one thought precludes thinking other thoughts, and providing a concrete answer to an attitude questionnaire forces one to reduce the fuzziness of one's thoughts to a single and concrete response. So, in our view appreciating the impact that psychological measurement might have on psychological processes has nothing to do with going full-physics, but simply means that we go full-psychology.

All in all, it seems at the very least worthwhile to explore the consequences of assuming that measurement affects attitudinal processes. From the perspective of the AE framework, the most important factor in measurement is the extent to which the measurement instrument directs attention to the attitude object. Measurement instruments limiting attention to the attitude object affect entropy-reduction less than measurement instruments not limiting attention to the attitude object.[1] All else being equal, implicit measures therefore tap attitudes in states higher in entropy than explicit measures. From this follows that scores on implicit measures must principally be less consistent and less stable than scores on explicit measures. While there is a large number of findings in the attitude literature on implicit measures, the probably most robust results on implicit measures are (1) that they show low internal consistency and low temporal stability (e.g., Bar-Anan & Nosek, 2014; Gawronski et al., 2017; Hofmann et al., 2005) and (2) that their value in predicting behavior is limited (e.g., Oswald et al., 2013). Any theory on what is measured by implicit measures should therefore first and foremost address these findings – other-

[1] There is also evidence that mere measurement of intentions leads to stronger correspondence between intentions and behavior (e.g., C. Wood et al., 2016). This effect is also in line with the AE framework's implication that measurement of attitudes affects attitudinal entropy.

wise it would be severely limited in its empirical grounding. It is our view that most theories on implicit measures fall short of this criterion and because of this our understanding of what is measured by implicit measures has remained elusive. Because the AE framework straightforwardly explains the most robust findings on implicit attitudes, the AE framework shows strong potential to illuminate one of the most researched and yet most elusive constructs of the last two decennia of attitude research – that of implicit attitudes.

Linking the measurement of attitudes to other established phenomena in the attitude literature, such as the mere thought effect, not only provides more understanding of what it means to measure attitudes implicitly, but also provides a host of novel predictions regarding the measurement of attitudes. For this reason, we appreciate it that Van Dessel et al. (2018) already tested one of these predictions – specifically, the prediction that reliability of implicit measures is expected to be higher for attitude objects individuals think frequently about, than for attitude objects individuals do not think often about. In a straightforward test, Van Dessel et al. provided first support for this prediction. The result of this test is, of course, not conclusive support for the measurement implications of the AE framework, but it shows that that the AE framework is able to advance our understanding of implicit measurement. We are therefore looking forward to more empirical investigations of the AE framework's measurement implications – in our view, the question to what extent measurement affects attitudinal processes is a rather open one and better understanding of this issue will lead to theoretical progress in the understanding of attitudes.

7.9 Extending the AE Framework to the Interpersonal Realm

While the former sections of this reply mostly focused on clarifying and discussing principles and implications of the AE framework, we also want to use the opportunity of this reply to discuss possible extensions of the AE framework. Specifically, we want to reflect on Rios and Roth's (2018) thoughtful analysis of how one could apply the AE framework to the interpersonal realm.

In our target article, we already raised the question to what extent the different levels of attitudinal entropy reduction are socially instigated. Based on Rios and Roth's (2018) commentary, the answer to this question seems to be quite a lot. In the AE framework, the influence of the social environment can take two routes. First, the social environment might directly affect the information on which the attitude is based (by affecting the thresholds in the attitude network). Such a situation would arise when one, for example, belongs to a homogeneous social network where most individuals have the same political views (cf., Visser & Mirabile, 2004). Second, the social environment might affect the motivation for attitudinal entropy reduction (by affecting the dependence parameter of the attitude network). Such a situation would arise when, for example, a given topic receives much attention in the media. Other situations pointed out by Rios and Roth (2018) that might affect entropy reduction are wanting to confirm to the group norm (cf., Carlson & Settle, 2016) and the motivation to not look like a hypocrite (cf., Barden, Rucker, & Petty, 2005; Barden, Rucker, Petty, & Rios, 2014). In our view, a potentially fruitful avenue for future research is to investigate how dependent the different levels of attitudinal entropy reduction are on the social environment. It seems likely that high entropy reduction is often instigated by the social environment, as, for example, the typical determinants of attitude importance, such as values and social identification (Boninger et al., 1995), have clear social connotations. In our view, an intriguing possibility for the determinants of attitudinal entropy reduction would be that paying attention and directing some thinking to the attitude object have the default effect of moderately reducing attitudinal entropy and that higher entropy reduction is mostly instigated by the social environment – the situations Rios and Roth discuss are obvious examples for this, but also having to make a decision (a likely core determinant of attitudinal entropy reduction) does generally not happen in a social vacuum.

Another question that we raised in our target article was what determines tolerance of attitudinal entropy. While it seems that attitudinal entropy reduction is a natural response in many situations, there are also situations that make attitudinal entropy reduction *less* likely. Rios and Roth (2018) also provide some ideas regarding these determinants of tolerance toward attitudinal entropy, such as being motivated to adopt different attitudes depending on the context (Sinclair, Lowery, Hardin, & Colangelo, 2005) and to integrate different viewpoints (Antonio et al., 2004). In our target article, we briefly discussed empirical evidence of cultural differences in attitudinal entropy reduction. An exciting possibility for cultural differences in attitudinal entropy reduction might be that the low levels of attitudinal entropy are more or less culturally invariant, while the higher levels are dependent on the cultural context and that the higher levels of attitudinal entropy reduction take different forms in different cultures.

For example, in individualistic cultures higher levels of attitudinal entropy reduction might be more determined by intrapersonal processes, while they might be more determined by interpersonal processes in collectivistic cultures (cf., Hoshino-Browne et al., 2005; Kitayama et al., 2004).

A final point that we want to make regarding the interpersonal consequences of attitudinal entropy reduction is that we wholeheartedly agree with Rios and Roth (2018) that attitudinal entropy reduction can have negative consequences (both on the individual level and on the group level). As Rios and Roth point out, high attitudinal entropy reduction in groups might be indicative of group think (Janis, 1982) and also on the individual level low entropy attitudes can have negative consequences (such as being too rigid and extreme). Better understanding the determinants of attitudinal entropy reduction might therefore also help in providing interventions to combat these negative consequences.

7.10 Attitudinal Entropy Framework – The App

While we believe its formalized nature is one of the AE framework's assets, we also acknowledge that most social psychologists are not familiar with working with formalized theories. To make the AE framework more accessible, we therefore implemented the basic workings in an online app using the interactive web application Shiny (Chang, 2018) that runs on the programming platform R (R Core Team, 2013). The app can be assessed under the url https://jdalege.shinyapps.io/AttitudinalEntropyFramework/ and we will keep updating the app so that most insights from the AE framework are eventually implemented in the app.

Currently, the app has two main tabs. In the first tab, the user can simulate the equilibrium distribution of a given network under two different dependence parameters, which are used to model different amounts of attention and thought directed at the attitude (similar to the simulations on the mere thought effect in the target article). Currently, the user can vary the size of the networks between 4 and 10 nodes, the values of all thresholds between -1 and 1, the connectivity of all edges in the network between 0, 0.05, and 0.1, and the dependence parameter between 0 and 3. Figure 7.2 shows a screen shot of this tab in which we replicated the first simulation of the target article on the mere thought effect (Simulation 2a). In the second tab, the user can simulate the dynamic behavior of the AE framework (similar to the simulations on persuasion in the target article and in the current reply). The starting thresholds are currently automatically set to 0.2 and the user can vary the amount (between 0 and 10) and

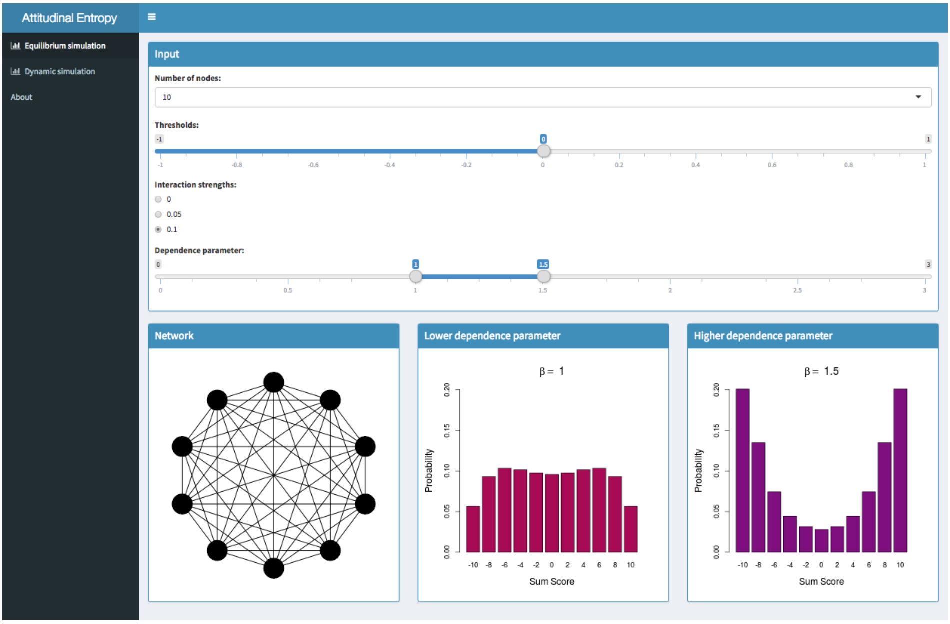


Figure 7.2: Screenshot of the online application in which Simulation 2a of our target article is replicated.

strength (between -1 and 0) of the thresholds that will be changed by the persuasion, the dependence parameter between 0 and 3, and the number of simulated individuals between 10 and 100 (note that a high number of individuals results in a rather long time of computation). Figure 7.3 shows a screen shot of this tab in which we replicated the low dependence parameter/3 weakly changed thresholds condition of Simulation R1.

7.11 An Essential Next Step

To conclude, we want to reiterate our claim that an essential next step for attitude research and psychology in general is to develop theories that (a) rely on few fundamental principles, (b) are formally specified and (c) make unambiguous empirical predictions. We remain, of course, open to the possibility that the specifics of the AE framework might be overthrown by a similarly well-specified theory, but as we show in this reply there are ample reasons to think that the AE framework rests on the correct building blocks to construct a comprehensive theory of attitudes. In our view, the main tasks for future research on the AE framework will revolve around two issues. First, the testing of predictions derived from the AE

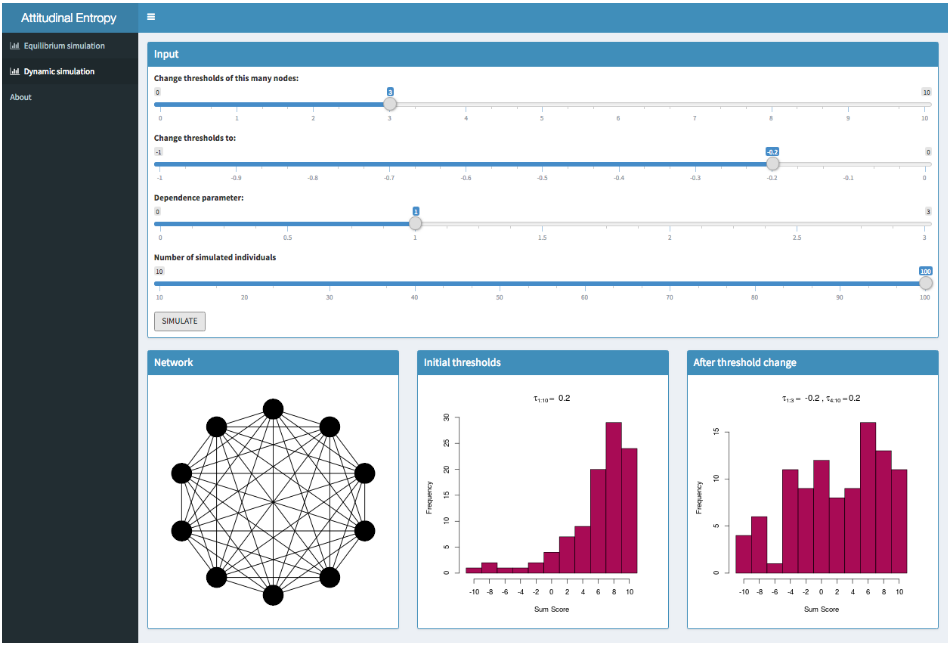


Figure 7.3: Screenshot of the online application in which the low dependence parameter/3 weakly changed thresholds condition of Simulation R1 is replicated.

framework to establish a firmer empirical basis for the AE framework and second, studying the boundary conditions of the AE framework. Based on the parsimonious nature of the AE framework it is certain that at some point the explanatory scope of the AE framework will break down and the explanatory power of the AE framework will be determined by how fast this point is reached. Given the many established findings the AE framework is able to explain, we are optimistic that this power is rather substantial. We therefore stand by our conclusion of the target article that the answer to the question why we think might be ”to reduce the entropy of our mental representations”. However, we also agree with March et al. (2018) that thinking is always in the service of doing, but the question remains how the cognitive system accomplishes to think in a functional way. Our tentative answer to this question is, however, straightforward and instructive: thinking is for reducing entropy, and reducing entropy is for doing.

Chapter 8

The Learning Ising Model of Attitude (LIMA): Entropy Reduction by Hebbian Learning

Abstract

This chapter introduces the Learning Ising Model of Attitude (LIMA). The LIMA integrates Hebbian learning into the Attitudinal Entropy (AE) framework, which represents a general theory on attitude built on statistical mechanics concepts and based on the central principle that attention and thought reduce attitudinal entropy (i.e., instability and unpredictability of attitudes). The AE framework uses the Ising model as formalization and attention and thought are modeled by a dependence parameter. Using simulations, we first show that Hebbian learning leads to networks that efficiently reduce entropy. Additionally, we model feedback between instability of attitudes and the dependence parameter, implement learning of dispositions of attitude elements, and provide an illustration of attitude change. We discuss the LIMA's contribution to the unification of research on attitudes, its relation to Hopfield network models and Boltzmann machines, and potential avenues for future research.

This chapter has been adapted from: Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2018). *The Learning Ising Model of Attitude (LIMA): Entropy reduction by Hebbian learning*. Manuscript in preperation.

8.1 Introduction

Attitude – the liking or disliking of an object – is one of the social sciences' central concepts and there are few concepts that have received as much attention from researchers as attitudes have. The focus on attitudes has produced a vast number of empirical findings and the time seems ripe to integrate these findings in a general and formalized framework. Recently, we provided a proof of principle that such integration is possible by using concepts from statistical mechanics to derive a formal theory of attitude – the Attitudinal Entropy (AE) framework (Dalege et al., 2018a). The AE framework rests on the central principle that attention and thought reduce attitudinal entropy (i.e., instability and unpredictability of attitudes) by increasing dependence of attitude elements (i.e., beliefs, feelings, and behaviors vis-á-vis an attitude object) connected in a network structure and encompasses several hallmark findings from the attitude literature. However, an important question that is not yet addressed by the AE framework is how attitude networks, which are able to reduce entropy, develop over time. While classic perspectives assume that attitudes are highly stable over time and would therefore represent attitudes low in entropy (i.e. attitudes can be 'readily' retrieved from memory; Fazio, 1995, 2007; Fazio et al., 1986), modern perspectives emphasize the dynamical nature of attitudes, that develop as a result of various contextual influences (Schwarz, 2007; Wilson & Hodges, 1992) and would therefore represent attitudes relatively high in entropy. The current chapter addresses these dynamics and we show that a basic learning mechanism, Hebbian learning, leads to attitude networks that efficiently reduce entropy and test implications of combining the AE framework with Hebbian learning. Combining the AE framework with Hebbian learning leads to the implication that efficient reduction of attitudinal entropy is only possible when two conditions are met – Hebbian learning must be present and at least a moderate amount of attention must be directed at the attitude object.

We thus currently build on the AE framework, in which attitudes are conceptualized as networks of interrelated elements that *ceteris paribus* gravitate towards inconsistency (high entropy). This is in line with the second law of thermodynamics, according to which entropy of any closed system *always* increases – implying that closed systems always become increasingly disordered and unpredictable. However, for attitudes to effectively fulfill their functions, such as to organize knowledge, increase utility, and express values (Katz, 1960; Smith et al., 1956), attitudes must be in low entropy states. More in general, for the human mind to function, it must reduce the entropy of its representations, such as attitudes (Cunnningham et al., 2013; Dalege et al., 2018a; Hirsh et al., 2012). The reduction of entropy might thus be of similar importance at the psychological level as it is on the biological level – the ability to reduce their internal entropy by increasing the entropy of their surroundings is the defining feature of living systems as Erwin Schrödinger (1944) argued in his seminal book *What is Life?*.

The AE framework specifies how the human mind achieves attitudes to be often times in low entropy states and in doing so builds on the Causal Attitude Network (CAN) model (Dalege et al., 2016), which conceptualizes attitudes as networks of attitude elements that have bidirectional influence on each other (e.g., judging a person to be honest makes it more likely that you judge the person also as caring and vice versa). The CAN model rests on psychometric network models as applied to cognitive psychology, psychopathology and personality psychology (e.g., Cramer et al., 2012, 2010; van der Maas et al., 2006) and constraint satisfaction models of attitude (e.g., Monroe & Read, 2008; Read et al., 1997; Shultz & Lepper, 1996; Simon, Snow, & Read, 2004; Simon et al., 2015). According to the CAN model, dynamics of attitudes can be described in an idealized fashion by the Ising (1925) model, which originated in statistical mechanics. Dynamics in the Ising model follow the principle of energy minimization and the energy of a configuration (i.e., states of the nodes in the network) is determined by the following equation:

$$H(\chi) = -\sum_{i} \tau_i \chi_i - \sum_{i,j} \omega_{ij} \chi_i \chi_j, \tag{8.1}$$

where $H(\chi)$ represents the Hamiltonian energy of a configuration of k distinct nodes (e.g., attitude elements) $1..., i, j, ...k$. χ represents the states of the nodes (-1, 1), which in combination with the nodes' thresholds (denoted by τ) and weights (denoted by ω) determine the Hamiltonian energy. The thresholds represent the disposition of the nodes to be in a given state, with positive (negative) values indicating that a given node has the disposition to be in a positive (negative) state. In attitude networks, these thresholds represent external information one has about the attitude elements (e.g., a positive threshold for judging a person as honest could indicate that one has often observed this person to act in an honest manner). The weights represent the strength of the (symmetric) interaction between nodes, with positive (negative) values indicating that two given nodes have an excitatory (inhibitory) interaction. In attitude networks, these weights represent influence between attitude elements (e.g., a strong positive weight between judging a person as honest and caring would indicate that these two beliefs are highly dependent on each other).

The likelihood of a given configuration at equilibrium is then given by the following equation:

$$\Pr(X = \chi) = \frac{exp(-\beta H(\chi))}{Z}, \tag{8.2}$$

where β represents the inverse temperature (in the AE framework and in the remainder of this chapter called dependence parameter) of the Ising model. High values of β result in low-energy configurations being more likely than high-energy configurations and by this β directly scales the (Gibbs) entropy of the system, because high likelihood of low-energy states results in an ordered and predictable system. Z is the sum of all possible configurations' energies and therefore ensures that the probabilities add up to 1.

The AE framework extends the basic idea of the CAN model that attitudes can be conceptualized as Ising models by tying fundamental principles of statistical mechanics to attitude processes. First, attitude inconsistency and instability are linked to entropy, as both concepts describe a disordered and unpredictable state of the system (i.e., the attitude). Second, energy in attitude networks represents a mechanism to locally evaluate the global entropy of the attitude. Third, inverse temperature (i.e., dependence parameter) represents the amount of attention and thought directed at the attitude object. Because of this, thought and attention have the effect that attitude elements become more dependent on each other and low-energy states become more likely, which reduces attitudinal entropy. A large number of established empirical findings were shown to follow from these principles (Dalege et al., 2018a). For example, the mere thought effect on attitude polarization (i.e., merely thinking about attitude objects leads to polarization; e.g., Tesser, 1978; Tesser & Conlee, 1975) directly follows when the dependence parameter of the attitude network is increased, low stability of implicitly measured attitudes follow from assuming that implicit measures (e.g., Implicit Association Test; Greenwald et al., 1998) tap attitudes in high entropy states, and systematic (heuristic) processing of arguments (e.g., Chaiken et al., 1989; Petty & Cacioppo, 1986) is more likely when the dependence parameter is high (low).

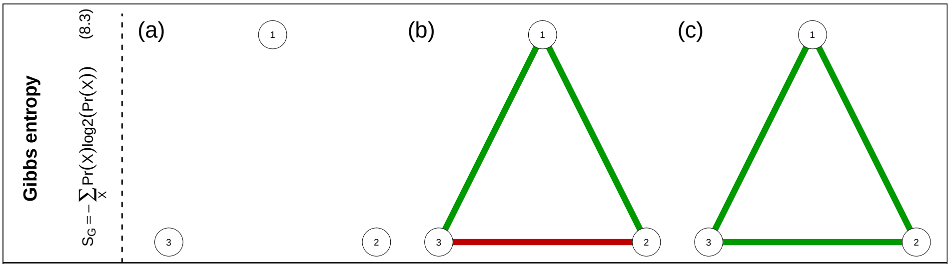
The findings supporting the AE framework, however, were obtained for a specific kind of networks – balanced networks. The idea of balanced networks dates back to the work by Fritz Heider (1946, 1958) focusing on social balance. Here a balanced network of three people would be that, for example, Harry and Mary are friends and both of them are also friends with John. In contrast, an unbalanced network would arise if one of these relations were negative (e.g., if John would dislike Mary). However, the AE framework is mute on why attitude networks should be balanced.[1]

The question why attitude networks would be balanced is crucial, because (a) there is growing evidence that attitude networks (and networks based on related constructs) are generally balanced (Dalege et al., 2016; Dalege, Borsboom, van Harreveld, Waldorp, & van der Maas, 2017; Dalege et al., 2018b; Dalege, Borsboom, van Harreveld, & van der Maas, 2017; Sachisthal et al., 2018; Sayans-Jiménez, van Harreveld, Dalege, & Rojas Tejada, 2018; Schlicht-Schmälzle, Chykina, & Schmälzle, 2018) and (b) only balanced attitude networks can efficiently reduce entropy (see Figure 8.1 for an illustration). Assuming neutral thresholds (i.e., all thresholds are set to 0) a network that is not balanced or unconnected cannot efficiently reduce its entropy. Note that we use Gibbs entropy as a measure of entropy throughout the chapter. As can be seen in Equation 8.3 in Figure 8.1, Gibbs entropy measures how likely different configurations of the system are. For example, Gibbs entropy in the network in Figure 8.1 would be at minimum when the only possible configuration would be that all nodes are in the state '1'(e.g., an attitude that is always entirely positive). In contrast, Gibbs entropy would be at maximum when all configurations are equally likely (e.g., an attitude that is completely unstable).

As can be seen in Figure 8.1a, in a completely unconnected network, the entropy of the network remains at maximum even when the dependence parameter (i.e., attention and thought directed at the attitude object) approaches infinity, implying that the attitude would remain highly unstable no matter how much attention and thought one directs at the attitude object. Similar behavior can be observed for an unbalanced network, in which the entropy only slightly decreases when the dependence parameter approaches infinity (see Figure 8.1b). In contrast, a balanced network results in decreased entropy even when the dependence parameter moderately increases and goes to low entropy when the dependence parameter approaches infinity (see Figure 8.1c), implying that some attention and thought would be sufficient to create a stable attitude.

A possible candidate to help explain the process through which bal-

[1] It is important to note here that balance in networks and entropy are related but distinct concepts. Balance concerns the edges in networks (e.g., whether individuals are friends in social networks or whether beliefs are dependent on each other in attitude networks) whereas entropy concerns states of the network's nodes (e.g., which beliefs individuals have in social networks or whether beliefs are endorsed or not in attitude networks). In general, balanced edges (e.g., everyone is friends with everyone) in networks lead to synchronized states of the nodes in the networks (e.g., everyone holds the same opinion).



$\beta = 0$	$S_G = 3$	$S_G = 3$	$S_G = 3$
$\beta = 1$	$S_G = 3$	$S_G = 2.62$	$S_G = 1.38$
$\beta \to \infty$	$S_G = 3$	$S_G = 2.58$	$S_G = 1$

Figure 8.1: Entropy reduction in unconnected, unbalanced and balanced networks. In (3) $Pr(X)$ refers to a given state's probability, (a) shows the values of entropy for an unconnected network under different dependence parameters, (b) shows the values of entropy for an unbalanced network under different dependence parameters and (c) shows the values of entropy for a balanced network under different dependence parameters. Green (red) edges represent positive (negative) ω. All $|\omega| = 1$ and all $\tau = 0$.

anced attitude networks develop is Hebbian learning. Hebbian learning in neural networks is based on the idea that "cells that fire together, wire together"; implying that if two nodes' states are correlated the connection between these nodes increases. There are several reasons suggesting that Hebbian learning also plays a role in the development of attitude networks. First, Hebbian learning is probably the simplest and widely used learning algorithm and relies only on associative learning without supervision. Second, connectionist models of attitude relying on Hebbian learning successfully reproduced several findings in the attitude literature (e.g., Monroe & Read, 2008). Third, there is some evidence that evaluative conditioning (i.e., the learning of associations between attitude objects and evaluations) follows Hebbian learning (e.g., Beckers, De Vicq, & Baeyens, 2009; De Houwer et al., 2001). The reason why Hebbian learning is likely to result in balanced attitude networks is that over the long run unbalanced triplets in a network such as the one shown in Figure 1b are likely to vanish because of differences in the edge weights. For example, if the negative edge in the network in Figure 8.1b were less strong than the positive edges, Hebbian learning would further decrease the magnitude of the negative edge while increasing the positive edges, thereby making the network increasingly balanced. Because of this, in the long run the negative edge would eventually become positive. On the other hand, if the negative edge is stronger than one of the positive edges, say the edge connecting node 1 and node 2, the positive edge between node 1 and node 2 would decrease, while the other edges would increase, thereby again making the network increasingly balanced.

In the following we introduce the Learning Ising Model of Attitude (LIMA), in which we add Hebbian learning to the AE framework. After setting up this model, we first test whether indeed Hebbian learning results in balanced networks, after which we investigate several scenarios under this model to investigate the implications of the LIMA.

8.2 The Learning Ising Model of Attitude (LIMA)

The LIMA rests on two central dynamical rules. The first rule governs the updating of a node's state (in the present context the valence of an attitude element) and the second rule governs the updating of edge weights (the connections between attitude elements).

8.2.1 State Updating

The state updating of the LIMA is governed by Glauber dynamics (Glauber, 1963). At each iteration a random node is picked and the energy of this node is calculated using the following equation, which is adapted from Equation 8.1:

$$E(\chi)_i = -\tau_i \chi_i - \sum_{i,j} \omega_{ij} \chi_i \chi_j, \tag{8.4}$$

where i represents the picked node. The energy of χ_i, $E(\chi)_i$, is then contrasted with the energy of the node's opposite state $-\chi_i$, and the node flips its state with the probability calculated by the following equation:

$$\Pr(\chi_i \rightarrow -\chi_i) = 1/(1 + exp(-\beta(E(\chi)_i - E(-\chi)_i)), \tag{8.5}$$

At each iteration the state of a given node is evaluated in how well its state confirms to the network structure. If the node does not confirm well to the network structure (e.g., a belief is not in line with other positively connected attitude elements) the likelihood that this node changes its state is relatively high. In contrast, if the node confirms well to the network structure (e.g., a belief is in line with other positively connected attitude elements) the likelihood that this node changes its state is relatively low.

8.2.2 Edge Weights Updating

The weight updating in the LIMA is governed by Hebbian learning (Hebb, 1949). After updating its state, weights between the picked node and nodes that are in the same (different) state increase (decrease). This weight updating is implemented by using the following equation:

$$\Delta\omega_{ij} = \epsilon(1 - |\omega_{i,j}|)\chi_i\chi_j - \lambda\omega_{i,j}. \tag{8.6}$$

The first term of Equation 8.6 is responsible for Hebbian learning and has a soft bound of 1 and ϵ represents the learning parameter. The second term of Equation 8.6 results in decay of edge weights so that edge weights are driven towards 0 when the behavior of the network is random, with λ representing the decay parameter. We set both ϵ and λ to .001 in the simulations, resulting in learning and decay that is neither too extreme nor too weak.

8.3 Simulations

In the following simulations, we first illustrate the basic workings of the LIMA. In Simulation 1 we illustrate how a LIMA network develops under a constant dependence parameter and whether indeed Hebbian learning leads to more balanced networks in the LIMA. In Simulation 2, we then vary the dependence parameter to show that because of Hebbian learning the attitude network can reduce entropy in an increasingly efficient way. Second, we extend the LIMA to also include feedback between instability of the attitude network and the dependence parameter. By doing so, we implement the AE framework's assumption that entropy of attitude networks is indirectly evaluated through the instability of the attitude network and also provide a formalization of attitude importance. In Simulation 3, we illustrate the effect of modeling feedback between instability and entropy. Third, we extend the LIMA to also include learning of thresholds following a similar process as the learning of weights (representing that endorsing an attitude element makes it increasingly likely that the attitude element is endorsed in the future). In Simulation 4, we illustrate the effect of learning thresholds in the context when one threshold is inconsistent with the rest of the thresholds. Finally, we combine all the extensions of the LIMA to model attitude change. In Simulation 5, we illustrate the effect of reducing attitude importance in combination with changing the information on which the attitude is based (by changing the thresholds of the attitude network).

8.3.1 Simulation 1: Development of an Initial Attitude Network

In Simulation 1 we investigated under what circumstances a balanced network develops. In Simulation 1a, we tested whether a slight disposition of nodes to be in the same state is sufficient for consistent edge weights to develop. We started the simulation with all edge weights set to 0, all thresholds set to .2, and kept β constant at 1. Note that these settings result only in a weak disposition of the nodes being in the same state (in fact the thresholds cause each two nodes to be in the same state only 51.9% of the time). This would represent a situation in which an individual for example observes that when a person displays one positive trait there is also a tendency for this person to display another positive trait. For example, the individual might observe that when a person is generally honest that this person is also generally caring and nice. We ran the simulation for 5,000 iterations and let the weights update according to Equation 8.6.

The results showed that weak but consistent thresholds are sufficient for a weakly connected but balanced network to develop. As can be seen in Figure 8.2a, the network first develops seemingly random edge weights with only a moderate tendency for edge weights being positive and that after more and more iterations the network becomes increasingly balanced. These results lead to the conclusion that Hebbian learning leads to a balanced network when there is only a weak tendency for nodes to be in the same state. This implies that true positive correlations between observations become increasingly exaggerated in the network's representation (e.g., one overestimates how honest a caring person is).

In Simulation 1b, we tested whether the LIMA is also able to differentiate between nodes that have a slight disposition to be in the same state (e.g., nodes with the same valence) and nodes that have a slight disposition to be in different states (e.g., nodes with different valence). We used the same simulation setup as in Simulation 1a, except that the thresholds of node 5-10 were set to -.2. This would represent a situation in which an individual observes that when a person has one positive trait there is also a tendency for this person to *not* have another negative trait. For example, the individual might observe that when a person is generally honest that this person is generally *not* selfish or mean.

The results showed that the LIMA is also able to differentiate between nodes that have a slight disposition to be in the same state and nodes that have a slight disposition be in different states. As can be seen in Figure 8.2b, at the end of the simulation almost all nodes with the same threshold are positively connected, while nodes with different thresholds are negatively connected – we thus again observed the development of a bal-

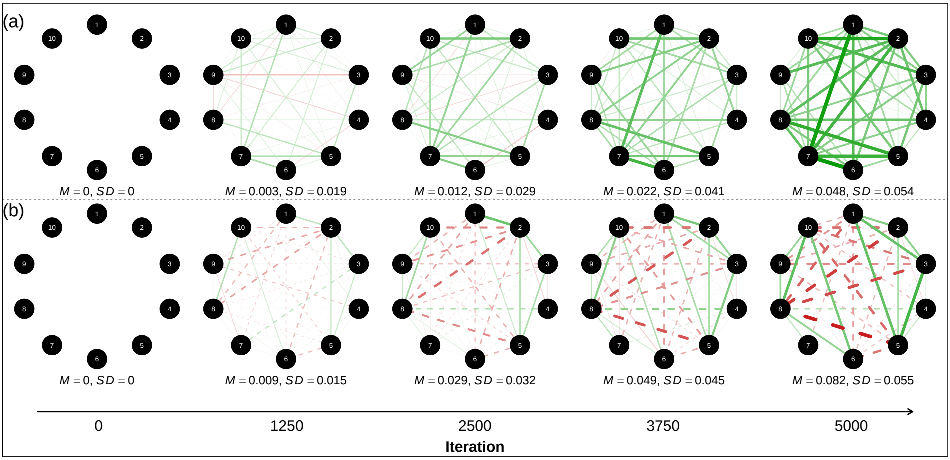


Figure 8.2: Development of Initial Attitude Network. (a) shows the development of the network with positive thresholds at different iterations. (b) shows the development of the network with mixed thresholds at different iterations. Note that edges between the different threshold clusters are dashed and that for calculating the mean of the edge weights, edge weights between the different clusters were multiplied by -1.

anced network. These results lead to the conclusion that Hebbian learning enables the LIMA to differentiate between nodes of different dispositions (e.g., differentiating between positive and negative beliefs). This implies that also true negative correlations between observations become increasingly exaggerated in the network's representation (e.g., one underestimates how selfish an honest person is).

It is also noteworthy that the observed development of attitude networks is only possible when the dependence parameter is not too low. This underscores that attitude networks, which efficiently reduce attitudinal entropy, only develop when sufficient attention is directed at the attitude object,

8.3.2 Simulation 2: Decrease of Entropy through Learning

In Simulation 2 we investigated the effect of learning on the effectiveness of entropy-reduction. We used a fully-connected ten-node network in which the initial edge weights were set to low positive values drawn from a normal distribution with M = .05 and SD = .01 (the same initial weights were used for all following simulations). Thresholds were all set to 0 (unless explicitly noted, the same holds for all following simulations). The dependence parameter was varied in the following way: In the first 400 iterations β increased linearly from 0 to 2, in the second 400 iterations

β decreased linearly to 0, and in the third 400 iterations β remained at 0. This procedure was repeated four times, resulting in a total of 4,800 iterations. Based on the variation in the dependence parameter we thus model that one directs varying amounts of attention and thought at the attitude object (e.g., one interacts with a person on different occasions). We ran this simulation twice: once without learning and once with learning of edge weights.

Figure 8.3 shows the main results of the simulation. As can be seen in Figure 8.3a, without Hebbian learning entropy-reduction remains ineffective (e.g., one would not be able to make up one's mind about a target person). In contrast, as Figure 8.3b shows, with Hebbian learning, entropy-reduction becomes more and more effective, because edge weights continuously increase (e.g., one has an increasingly certain attitude about the person). As can be seen in the first cycle of increasing the dependence parameter, the Gibbs entropy only slightly decreases while in later cycles the Gibbs entropy approaches the minimum value of 1 (Gibbs entropy of 0 would only be possible with non-zero thresholds). It is also noteworthy that with increase in the edge weights the network approaches low entropy states more effectively and also stays in low entropy states for a longer time. The LIMA is thus able to decrease its Gibbs entropy without directly evaluating this measure. These simulations also again illustrate that efficient attitudinal entropy reduction can only be accomplished with the combination of attention and Hebbian learning

8.3.3 Simulation 3: Coupling of Instability and Attention

Simulations 1 and 2 illustrate the basic workings of the LIMA and the following simulations focus on extensions of the LIMA. First, based on the AE framework we implement feedback between instability of the attitude network and the dependence parameter – high instability leads to a higher dependence parameter. The AE framework holds that instability of attitudes results in feelings of ambivalence and this in turn leads to an increase in attention and thought directed at the attitude object (cf. van Harreveld et al., 2015), which then reduces the attitude's instability and indirectly the attitude's entropy. The reason that the AE framework assumes that entropy is indirectly evaluated through the instability of an attitude is that evaluating instability is computationally much more efficient than directly evaluating entropy. The assumption that instability of attitudes influences feelings of ambivalence is indirectly supported by the finding that physical signs of instability (e.g., moving from side to side) heighten felt ambivalence (Schneider et al., 2013). There is also in-

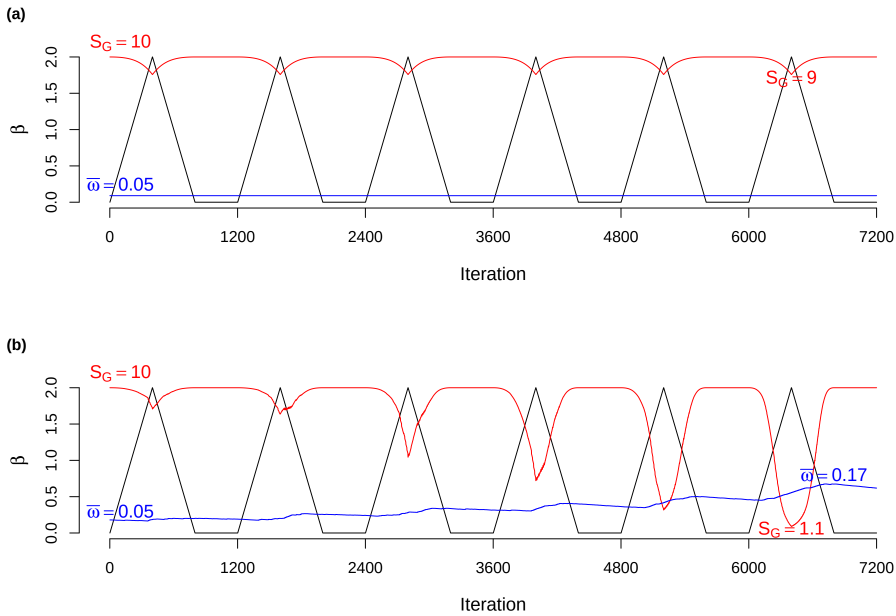


Figure 8.3: Effect of learning under varying dependence parameters. (a) shows the development of the Gibbs entropy (red line) and edge weights (blue line) under varying dependence parameter (black line) without Hebbian learning of edge weights. (b) shows the effect of Hebbian learning of edge weights.

direct evidence that individuals are aware of their attitude instability, as individuals have knowledge about their attitude clarity (i.e., individuals know how certain they are what their 'true' attitude is) and this attitude clarity is also a strong predictor of subjective ambivalence (Petrocelli et al., 2007).

In addition to modeling feedback between instability and the dependence parameter, we also implement attitude importance as the equilibrium state of the dependence parameter (e.g., if an individual attaches high importance to an attitude object, the individual will continue to direct high levels of thought and attention at the attitude object). Modeling attitude importance in this way is based on the finding that attitude importance is strongly related to the amount of thought an individual directs at an attitude object (Krosnick et al., 1993).

To couple the attitude's instability and the amount of attention directed at the attitude object, we use the following equation:

$$\beta_{t+1} = (1-\zeta)\beta_t + \zeta(\iota + \iota I - \beta_t), \tag{8.7}$$

where ζ represents the strength of the coupling between stability and attention, ι determines the equilibrium state of β (the equilibrium state of β is equal to $\iota/2$) and can therefore be seen as a formalization of importance attached to the attitude object, and I represents the instability of the system and is evaluated in the following way:

$$I = \frac{1}{i} \sum_{t-i,t} C, \tag{8.8}$$

where C represents whether the picked node changed its state in a given iteration and i determines how many of the preceding iterations are evaluated.

In Simulation 3 (and all following simulations), we set ζ to .001 and i to 10, because these values result in smooth changes of β, while evaluating ten iterations is also easily computable. We varied the importance parameter ι between 1, 2, and 3. Varying the importance parameter in such a way represents processing of an attitude object, to which one attaches little importance (e.g., a colleague one seldom interacts with), moderate importance (e.g., a friend one sees relatively often), and high importance (e.g., one's partner). At the beginning of the simulation β was set to 0 (representing that initially no attention and thought is directed at the attitude object; e.g., one does not know the person yet) and we let the simulation run for 6,000 iterations.

The main results of Simulation 3 can be seen in Figure 8.4. As can be seen in Figure 8.4a when the importance parameter ι is not sufficiently high the increase in β is not consequential – both instability and entropy remain high (e.g., one remains indifferent about the colleague). Figure 8.4b and 8.4c, in contrast, show that with sufficiently high importance parameter ι stability is increased and through this entropy is reduced and the strength of ι determines how quickly the low entropy state is reached (e.g., one has a certain attitude toward one's friend and partner). Also noteworthy is that when ι is sufficiently high the dynamic of β is simply that it quickly approaches its equilibrium value (see Figure 8.4c; e.g., one falls in love very quickly). When ι is lower, β either remains above its equilibrium level (see Figure 8.4a; e.g., one is not motivated to make up one's mind about the colleague) or rises above its equilibrium level to quickly drop back to its equilibrium level (see Figure 8.4b; e.g., it takes some time before one has a certain attitude toward a new friend). Finally, Simulation 3 also confirms that instability represents an indirect measure of entropy and therefore represents a computationally feasible way for the human mind to evaluate its attitudinal entropy.

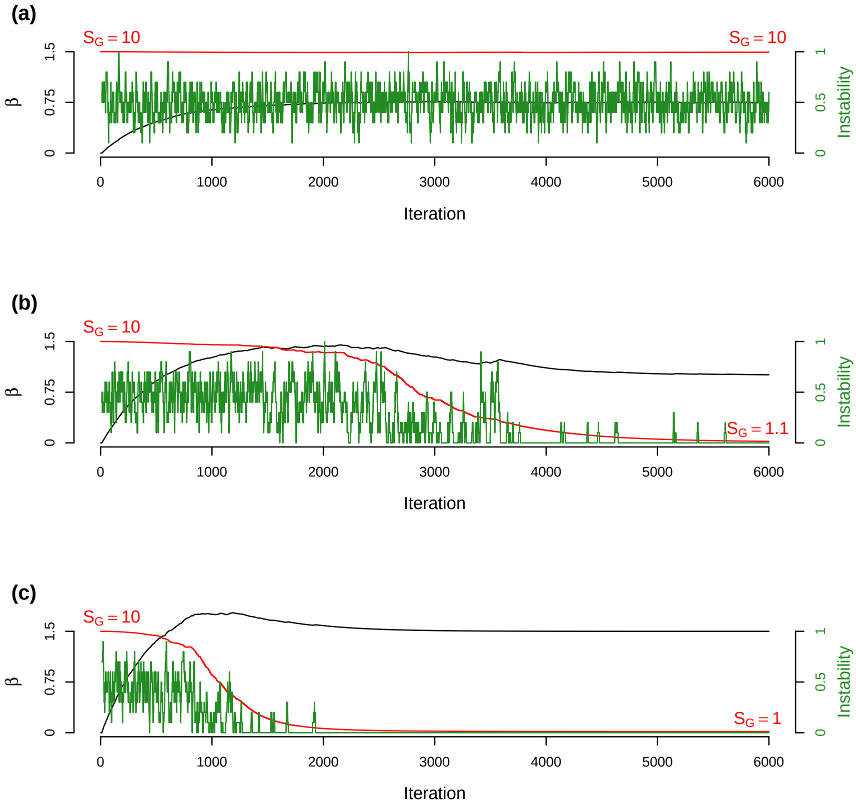


Figure 8.4: Effect of coupling instability and attention under varying importance parameters. (a) shows the development of β (black line), instability (green line) and the Gibbs entropy (red line) for $\iota = 1$. (b) shows these measures for $\iota = 2$ and (c) shows these measures for $\iota = 3$.

8.3.4 Simulation 4: Learning of Thresholds

In the following simulation we add a learning mechanism for the thresholds to the LIMA. This learning mechanism is analogous to the learning mechanism of the weights, implying that a consistently positive (negative) state of a given node results in an increasingly positive (negative) threshold of this given node. The learning of thresholds therefore models that endorsing an attitude element makes it increasingly likely that the element is endorsed in the future (e.g., when an individual judges a person frequently as honest, they are likely to become increasingly inclined to continue judging this person as honest; Dalege et al., 2016; Fazio, 1995).

Implementing this mechanism results in thresholds being not simply exogenous to the model but that the development of thresholds is (partly) accounted for by the model. The information that the attitude relies on therefore does not reflect a property that is solely dependent on the environment but also reflects interpretations of the information provided by the environment.

To implement threshold updating, we used the following equation, which mirrors Equation 8.6, responsible for the weight updating:

$$\Delta\tau_i = \epsilon(1 - |\tau_i|)\chi_i - \lambda\tau_i. \tag{8.9}$$

As in Equation 8.6 ϵ represents the learning parameter and λ represents the decay parameter. In the following simulation, we set both these parameters to .001, which is the same value as for the weight updating.

Threshold updating is most relevant in the situation in which thresholds are incongruent. Such a situation would, for example, arise when an individual has negative beliefs and feelings toward a behavior, but is not able to quit the behavior, because of a strong habit or addiction (e.g., an individual, who wants to stop smoking but is not able to do so). In Simulation 4 we therefore modeled such a situation by setting the first threshold initially to a moderately strong positive value of .2 and all other thresholds to initial weak negative values of -.05. For this simulation, we set β to 1 throughout the simulation and we let each simulation run for 10,000 iterations. Because we observed qualitatively different results for different random initializations we ran 100 simulations and report the overall results of these simulations.

We found three qualitatively different results. First, the most common result was that the weights and thresholds developed in a way that a stable consistently negative attitude emerges (66% of the simulations; see Figure 8.5a for an example). In this case, the threshold of the first node changed to being negative, while all other thresholds increased in their negativity and all edge weights became increasingly positive. This would thus represent a situation in which the individual managed to quit the disliked behavior. Second, another common result was that the whole attitude flipped and a stable consistently positive attitude emerged (24%; see Figure 8.5b for an example). In this case, the threshold of the first node remained negative, all other thresholds also became negative, and all edge weights became increasingly positive. This would thus represent a situation in which the individual was not able to quit the disliked behavior and then changed her beliefs accordingly (a situation reminiscent of cognitive dissonance, Festinger, 1957). The final relatively common result was that an inconsistent network emerged (10%; see Figure 8.5c for

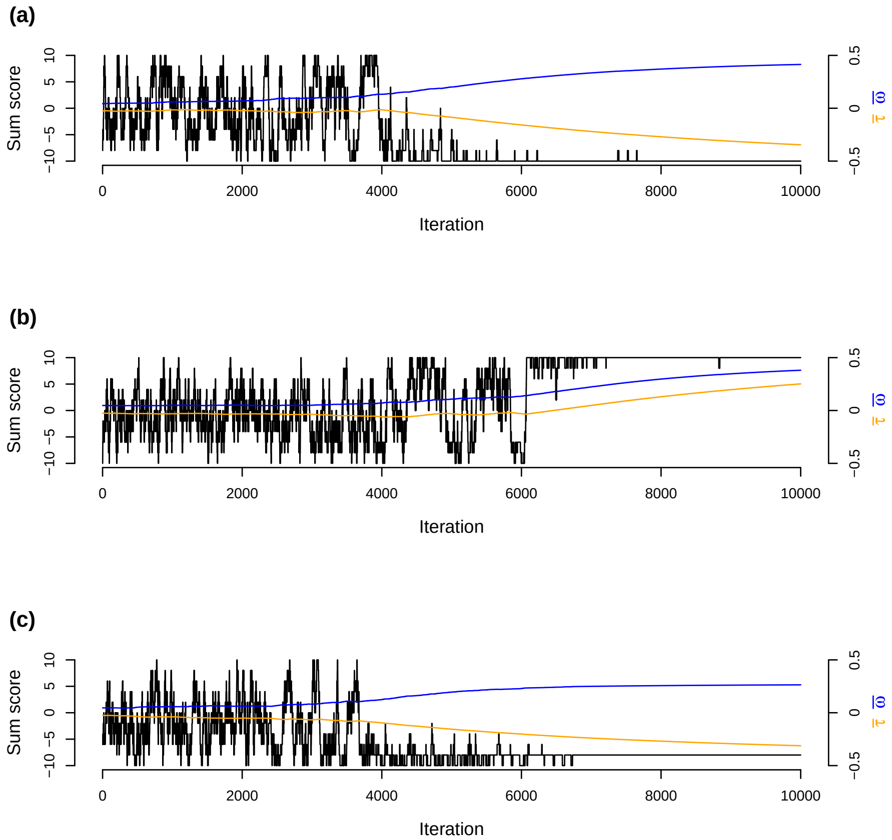


Figure 8.5: Effect of updating thresholds. (a) shows the development of the sum score of the attitude network (black line), mean weights (blue line) and mean thresholds (orange line) for three different exemplars of the typical results of Simulation 4.

an example). Typical for this result was that no threshold changed its direction, but all thresholds increased in their value. As a consequence of this, node 1 became negatively connected to all other nodes. This would thus represent a situation in which the individual was not able to quit the disliked behavior and developed inconsistent connections between beliefs and the behavior.

8.3.5 Simulation 5: Attitude Change

In the following simulation, we combine the extensions of the LIMA to model one of the most important questions in the attitude literature – under what circumstances do attitudes change? The AE framework holds that attitude change depends on two factors (Dalege et al., 2018a). First, it depends on changes in the thresholds of the attitude elements (e.g., by providing information regarding the attitude object). Second, change depends on the strength of the dependence parameter – the higher the dependence parameter the less likely the attitude is to change (e.g., highly important attitudes are resistant to change; cf. Howe & Krosnick, 2017).

In Simulation 5, we tested whether reducing the dependence parameter indeed results in more attitude change. For this, we first initialized the attitude network with a strongly positive disposition by setting all thresholds to .2 and we set the importance parameter ι to 3, so that the equilibrium value of β was equal to 1.5 (representing a highly important attitude that has a strong positive disposition; e.g., a highly involved supporter of a politician). We then let the network develop for 1,200 iterations. Then we set all thresholds to either -.2, representing strong persuasion (e.g., providing a lot of evidence that the politician is not suited for office), or we set all thresholds to -.05, representing moderate persuasion (e.g., providing some evidence that the politician is not suited for office). Additionally, the importance parameter ι remained either at 3, representing no change in the importance of the attitude, or changed to 1, representing strong decreases in the importance of the attitude (e.g., convincing the person that it is not consequential what she thinks of the politician). We then let the network develop for another 1,200 iterations. We repeated every combination of threshold change and importance parameter change 50 times.

The results indicated that attitude change depends to a large extent on the change in the importance parameter ι, see Figure 8.6. With a lowered importance parameter, attitude change is likely even when the persuasion was of moderate strength. In contrast, when the importance parameter remains high, attitude change is unlikely even when the persuasion attempt was strong. These results imply that attitude change can be accomplished more effectively when importance of the attitude is lowered. Changing the attitude of highly involved individuals is thus predicted to be more effective when one first lowers their involvement.

8.4 Discussion

William James (1890, p. 960) famously wrote "thinking is first and last for the sake of... doing", but the question has remained how thinking supports doing. To be able to act on mental concepts, such as attitudes, these concepts have to be in low entropy states. The present research

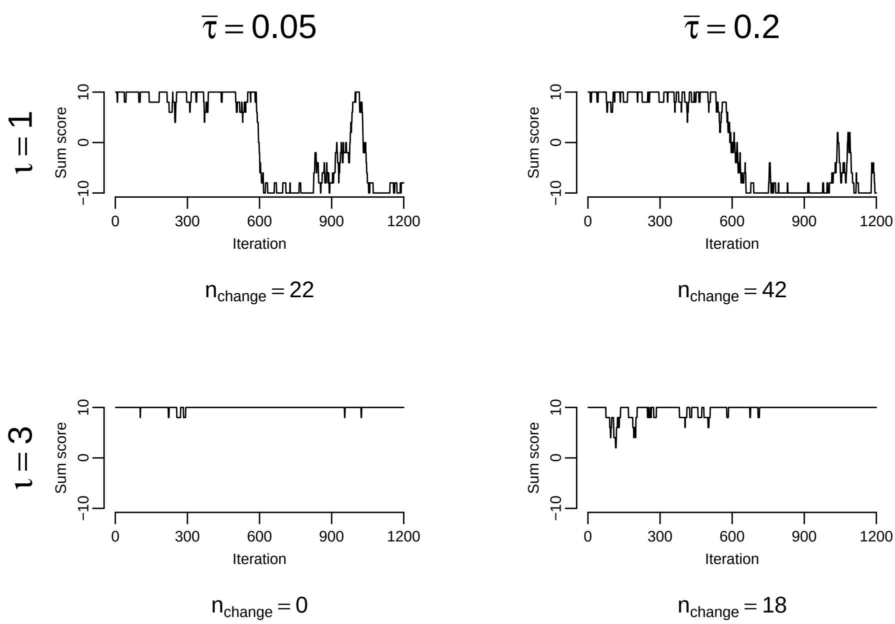


Figure 8.6: Effect of threshold change and importance parameter on attitude change. The four plots are each a typical example of the dynamics of the sum score of the network given the specified parameters.

shows that networks, which efficiently reduce entropy, straightforwardly develop when edge weights update according to Hebbian learning and therefore provides a possible explanation how the human mind reduces its entropy and by this creates mental representations on which it can act. The simulations in this chapter suggest that such mental representations can only be formed through the combination of attention directed at the attitude object and Hebbian learning of connections between attitude elements.

Using the LIMA, it becomes possible to integrate two seemingly contradictory views on attitudes. On the hand, classic perspectives on attitudes assuming attitudes are stable representations in memory (Fazio, 1995, 2007) would represent attitude networks with a long learning history (and therefore a balanced structure) that are high in dependence. Such attitude networks would appear to be attitudes stored in memory, because the strong dependence between attitude elements would hold them in the same state. On the other hand, more recent perspectives assuming attitudes are constructed on the spot (Schwarz, 2007) would represent attitude networks with either a short learning history and/or attitude networks that are low in dependence. Such attitude networks would be highly context-dependent and fluctuating. By integrating these two prominent (and thus far conflicting) views into a single framework, the LIMA contributes to the unification of research on attitudes.

Another advancement that the LIMA constitutes for the study of attitudes lies in its focus on temporal dynamics. While the underlying assumption in attitude research is always that attitudes are dynamic and continuously develop as a result of internal motivational factors and the interaction with information and others providing this information (e.g., van Harreveld, van der Pligt, & de Liver, 2009), the vast majority of empirical studies on factors influencing attitudes do not take into account such dynamical processes. Also, the few attitude theories that do explicitly focus on the change of attitudes over time (e.g., dissonance theory, Festinger, 1957) do not formalize the process through which such change occurs. The LIMA aims to provide a framework that provides insight into how and when attitudes change over time as a result of persuading factors.

The LIMA is, of course, not the first network model that uses Hebbian learning for the updating of edge weights. Two broad classes of neural network models – Hopfield (1982, 1984) networks and Boltzmann machines (Ackley, Hinton, & Sejnowski, 1985) also use Hebbian learning. One important difference between these models and the LIMA, however is that the dependence parameter (or temperature) has a fundamentally different meaning in both Hopfield networks and Boltzmann machines than in the LIMA. While in the LIMA the dependence parameter straightforwardly models attention and thought and thus creates a connection to entropy reduction, (1) in Hopfield networks the dependence parameter is set to infinity (temperature is set to 0), resulting in a completely deterministic system. Therefore, in Hopfield models there is no direct implementation of varying entropy of mental representations. (2) In Boltzmann machines the dependence parameter starts out at a low value and then increases to a higher value so that the system does not end up in a local optimum. Therefore, in Boltzmann machines the dependence parameter has no theoretical meaning and is only used as a tool for more efficient learning. Thus, in contrast to Hopfield models and Boltzmann machines, the LIMA creates a direct theoretical connection between Hebbian learning and entropy reduction and therefore represents a straightforward choice to model mental entropy.

Given the similarity between Hopfield models and the LIMA, the LIMA also provides an opportunity for integration between neural network models of attitudes that are based on Hopfield models, such as the Attitudes as Constraint Satisfaction (ACS) model (Monroe & Read, 2008) and the Attitudinal Entropy (AE) framework. An interesting avenue for future research would be to investigate whether more complicated models like the ACS model make predictions that cannot be derived from the more parsimonious LIMA. Such predictions can then be used to investigate how much complexity is needed to model the full range of established findings in the attitude literature.

By connecting psychometric network models (e.g., Epskamp, Maris, et al., 2018; van Borkulo et al., 2014) to Hebbian learning, the LIMA also provides a more general connection between the two broad approaches of neural network models and psychometric network models. Crucially, by doing so the LIMA provides an explanation for the finding that psychometric networks are generally rather densely connected. For example, in depression networks most symptoms are connected to each other (Cramer et al., 2016; Fried, Epskamp, Nesse, Tuerlinckx, & Borsboom, 2016; van Borkulo et al., 2015). At least some of the connections might arise due to Hebbian learning, which would also fit the finding that the risk of reoccurrence of a depressive episode becomes increasingly pronounced the more episodes the individual already experienced (Burcusa & Iacono, 2007).

It is our view that the LIMA not only provides a theoretical framework for attitudes, but that it might also illuminate other psychological processes, such as creativity and cognitive development. Creative solutions generally involve making connections between concepts that seem unrelated (Koestler, 1964; Simonton, 1999). From the perspective of the LIMA, such connections are most likely to arise when the dependence parameter is relatively low. This implies that creative solutions generally emerge when there is no strong pressure to come up with a solution, because otherwise the dependence parameter might increase to early. This theorizing relates to findings that promotion focus (i.e., focusing on obtaining positive outcomes rather than on avoiding negative ones; Higgins, 1997) and distraction increase creativity (Baas & De Dreu, 2008; Baird et al., 2012).

Regarding cognitive development, the LIMA might be applied to transitions between developmental stages (Piaget & Inhelder, 1969). From the perspective of the LIMA, the transition between two developmental stages reflects drastic changes in the connections between concepts. Such a situation would arise if one were to realize that current connections between concepts are inaccurate. As an example, take the balance scale task, in which children predict movement of a balance scale on which weights of different heaviness are placed at different positions (Inhelder & Piaget, 1958). To solve this task, children apply different rules that differ in their accurateness (Siegler, Strauss, & Levin, 1981). The simplest rule is that only the heaviness of the weights determines the movement of the scale, implying that there is a strong connection between heaviness and movement in the child's mind. Because this rule only results in the right answer in some instances, a child might learn that this connection is too strong and that other connections also have to be learned (e.g., taking also distance of weights into account). Such a process might be integrated in the LIMA by introducing feedback between the dependence parameter and accuracy of a connection, so that the dependence parameter drops when accuracy is too low to allow for the reordering of edges.

To conclude, the LIMA holds promise to integrate several disparate fields within and beyond the attitude literature and paves the way for several avenues for future research. Ultimately, the LIMA might not only represent a next step in developing a formal and general theory of attitude, but might also provide a further step in formalizing the main function of human thought and learning by tying these processes to the defining physical feature of living systems – the reduction of entropy.

Part IV

Conlusion

Chapter 9

General Discussion

9.1 Introduction

Is it possible to capture the immense complexity of the evaluative systems known as attitudes by few simple principles? In this dissertation I have shown that, if one uses the right lens to derive formalisms on attitude, the answer to this question might be ”yes”. The starting point of the approach taken in this dissertation is that attitude elements (i.e., beliefs, feelings, and behaviors vis-á-vis an attitude object) causally interact with each other. Based on this principle I first derived a measurement model of attitude – the Causal Attitude Network (CAN) model – and then extended this model to a general theory of attitude – the Attitudinal Entropy (AE) framework – by assuming that directing attention and thought at the attitude object increases the interactions between attitude elements and thereby decreases attitudinal entropy. Additionally, I provided the first empirical tests of the CAN model, showed that the AE framework can explain a host of established findings in the attitude literature, and developed a model that adds a learning mechanism to the AE framework. In this General Discussion, I want to take a step back and first discuss how the different formalizations developed in this dissertation relate to each other and in which situations one needs to focus on one or the other. Second, I discuss some limitations and issues that need to be addressed to further develop the approach formulated in this dissertation. Third, I place this dissertation’s approach in a broader context by discussing its implications for the nature of attitude, show that the approach provides a novel perspective on psychological measurement and mental causation, and connect the approach to recent developments in philosophy of science.

9.2 CAN Model, AE Framework, and LIMA – What Is Needed to Explain What?

While the different formalizations developed in this dissertation all depend on each other, there are also crucial differences between them. First, the Attitudinal Entropy (AE) framework can be seen as the overarching theory for the formalizations in this dissertation. Because of the limited number and abstractness of its principles, it is mostly concerned with general laws and has a strong nomothetic flavor. This is also illustrated by the fact that we used the simplest network imaginable (i.e., a network in which all nodes are equally connected) to model the AE framework's implications. Strikingly, despite the fact that we used such a highly simplified abstraction of attitudinal processes, we were able to reproduce several established findings from the attitude literature and thereby provide support for the assumption that many attitude processes follow relatively simple laws.

There are, however, certainly situations in which processes in addition to the ones specified by the AE framework have to be modeled to account for some findings. For example, to model different attitude bases (e.g., affective versus cognitive bases of attitude; Edwards, 1990; Fabrigar & Petty, 1999) one needs to also investigate the structure of attitude networks in addition to their general principles. This is where the Causal Attitude Network (CAN) model comes into play. By considering that attitude elements differ in their structural importance also more nuanced findings from the attitude literature can be modeled and as I have shown in Chapter 5, specific predictions regarding the impact of attitude elements on behavior can be made. Furthermore, in contrast to the AE framework, the CAN model can be directly applied to empirical data and can therefore provide more specific predictions of how behavior relates to attitude networks. For example, one could first estimate the attitude network structure of a group of individuals and then (by applying the principle of the AE framework that attention and thought increase dependency between attitude elements) one could make specific predictions of the consequences of this principle for the given attitude network structure (e.g., if the network is only weakly connected, directing attention and thought at the attitude object will only have a weak effect). Additionally, the CAN model can easily be extended to integrate factors related to attitude elements, such as social norms, perceived behavioral control and intentions, as specified by the theory of reasoned action (Fishbein & Ajzen, 1975) and its successor the theory of planned behavior (Ajzen, 1991).

Both the CAN model and the AE framework focus on already developed attitude networks and it is an interesting observation that, even without accounting for the development of attitude networks, many findings in the attitude literature can be explained through the AE framework. The reason for this is probably that experimental studies with short duration are the typical paradigm in the attitude literature. However, a comprehensive theory of attitude also needs to take the oftentimes longer life cycle of attitudes into account. Because of this I also specified a learning mechanism for attitude networks and implemented this mechanism in the Learning Ising Model of Attitude (LIMA). In contrast to the CAN model and the AE framework, the LIMA provides a framework to investigate attitude processes on a longer time scale. Crucially, it also provides a straightforward explanation for why attitude networks differ in their connectivity. While the AE framework specifies which situations increase the dependency between attitude elements and thereby increase attitude strength temporarily, the LIMA thus specifies how lasting differences in attitude strength develop over time.

A final note of clarification regards the difference between connectivity and the dependence parameter as specified in these formalizations. Because the CAN model is a measurement model, it only estimates connectivity of attitude networks; the dependence parameter is not estimated. The reason for this is that estimating both connectivity and the dependence parameter results in an unidentifiable statistical model (Epskamp, Maris, et al., 2018). Therefore, estimated connectivity (as in Part 3 of this dissertation) reflects both true connectivity as well as the dependence parameter. At a theoretical level, however, there is an important difference between connectivity and the dependence parameter: The dependence parameter varies depending on situational variables (e.g., having to make a decision) as specified by the AE framework, whereas connectivity depends on the learning history of the attitude as specified by the LIMA.

9.3 Limitations and Extensions

The approach formulated in this dissertation is able to explain a host of findings in the attitude literature, but several questions remain to be addressed in order to further develop the approach. These questions fall into three different categories. First, the approach currently does not take broader connections of attitude elements into account and at least two extensions are worthwhile exploring – how do different attitude networks connect to each other within the person and how do attitude elements connect between persons? Second, the temporal dynamics need to be further investigated – how fast does the dependence parameter increase or decrease and how fast does learning of connections take place? Third, there are some methodological issues that need to be solved – how can we estimate individual attitude networks?

9.3.1 Intra- and Interpersonal Extensions

The formalisms developed in this dissertation focus on single attitudes – attitudes, however, are not isolated systems, but interact with different attitudes within the same person and with attitudes of different persons. For example, political ideologies can be represented as a set of more specific attitudes (e.g., Brandt, Sibley, & Osborne, 2019; Gerring, 1997) and individuals belonging to the same social circle generally hold similar attitudes (e.g., Goel, Mason, & Watts, 2010; McPherson, Smith-Lovin, & Cook, 2001). Adding intra- and interpersonal extensions to the approach formulated in this dissertation requires a broader level of abstraction. When the focus lies on single attitudes, microstates represent attitude elements, but when the focus lies on how these attitudes are connected intra- and interpersonally the attitude itself becomes part of the microstate consisting of several attitudes. Interestingly, there is some evidence that the formalisms developed in this dissertation also apply to the interaction between attitudes forming ideologies and to the interaction of attitudes of different individuals (e.g., Brandt et al., 2019; Galam et al., 1982; Galesic & Stein, 2019). It thus seems promising to connect these different level models and a crucial step in doing so would be to formalize how attitude elements are summarized into a global evaluation. To do so, one would also need to specify how the influence between global attitudes also extends to influencing lower level attitude elements.

Another interesting avenue for modeling the interactions between attitudes at both the intra- and interpersonal level would be that attitudes not only influence each other in terms of valence, but that dependence also spreads from one attitude network to the other. At the intrapersonal level this would imply that, for example, one's involvement regarding one's attitude toward one political issue influences one's involvement regarding another connected political issue. At the interpersonal level, in contrast, this would imply that, for example, the involvement of one's friends influences one's own involvement. Future research can focus on disentangling the different consequences and implications of these theoretical questions.

9.3.2 Temporal Dynamics

Dynamics unfold over time and the extent to which dynamics are fast or slow has important implications for the system's behavior – however, currently, the formalisms developed in this dissertation take time only indirectly into account. Two dynamics that beg the question on what time scales they operate are the lowering and heightening of the dependence parameter and the learning of edge weights. The formalisms developed in this dissertation implicitly assume that variations in the dependence parameter take place at relatively short timeframes (based on the fact that we modeled empirical findings of short-lived experiments using variations in the dependence parameter) and that learning of edge weights takes place at a relatively long timeframe (based on the fact that Hebbian learning needs several learning instances to heighten edge weights). Regarding variations in the dependence parameter, empirical findings suggest that these variations take place at the time-frame of few minutes, because, for example, the mere thought effect is generally investigated in the timeframe of letting individuals think for a few minutes about the attitude object (e.g., Tesser & Conlee, 1975). Additionally, research on how involvement determines heuristic versus systematic persuasion indicates that variations in the dependence parameter can happen almost instantly, because immediately after individuals learn that the topic is relevant to them they are more likely to process arguments in a systematic fashion (e.g., Petty & Cacioppo, 1984; Petty et al., 1981, 1983). However, we currently have little knowledge whether individual differences exist in how fast the dependence parameter increases or decreases and which variables might predict these individual differences. Having more accurate knowledge about these parameters would lead to more accurate predictions regarding, for example, how pronounced the mere thought effect would be for different individuals.

Similar questions can be asked about the time-scale of learning connections between attitude elements. In contrast to the dependence parameter, learning processes that change connections between attitude elements must take place at a relatively slow pace. The nature of Hebbian learning implies a relatively slow updating of edge weights, because it depends on learning of correlations in the environment and for observing robust correlations one must observe several instances in which two things co-occur (e.g., one would need to observe persons several times acting in an honest *and* caring way). If Hebbian learning were taking place in a fast fashion edge weights could not develop in a robust way (e.g., if the edge weights increased based on only one observation extreme fluctuations in edge weights would be the result). However, making the time-scale of learning connections more specific is currently not possible and we also have no knowledge of individual differences in the speed of learning connections. The question on what time-scale learning of connections takes place is relevant, because answering it would lead to a higher accuracy for predicting individual's attitude network connectivity and thereby also lead to higher accuracy in predicting the effects of increasing the dependence parameter (e.g., predicting which individuals show a mere thought effect). Fortunately, investigating the time-scale of learning connections is relatively straightforward and could be accomplished by exposing individuals to a novel attitude object for several days and then comparing effects of heightening the dependence parameter of this group to a control group. For example, the strength of difference in the mere thought effect could then be used as a proxy for the strength of learning.

9.3.3 Methodological Developments

One of the main aims of the approach formulated in this dissertation was to provide formalisms that have a close connection to measurement and estimation – while the approach accomplishes this aim in many instances (e.g., it provides a measurement model of attitude), there are some instances in which further methodological developments are necessary to achieve this central aim. An important advance would be the possibility to estimate individual attitude networks. While there are possibilities to estimate individual networks based on time-series data (e.g., Bringmann et al., 2013; Epskamp, Waldorp, Mõttus, & Borsboom, 2018) and by directly asking individuals to report their network connections (Frewen, Schmittmann, Bringmann, & Borsboom, 2013), both possibilities face serious challenges in the estimation of attitude networks. First, time-series data relies on intraindividual variation. However, at least in the case of highly connected attitude networks, variation at the individual level is likely very limited (e.g., my belief regarding Donald Trump's honesty shows close to no variation) and because of this the most highly connected attitude networks might not be estimated accurately. Another issue in estimating attitude networks based on time-series data is that it is also rather unlikely that change in one attitude element only influences other attitude elements hours later (which is the typical timeframe for time-series methods, such as event-sampling methodology; e.g., Reis & Gable, 2000), so effects are unlikely to show in the time-lagged estimates. However, recent methods on network estimation on time-series data are also able to estimate a contemporaneous network (Epskamp, Waldorp, et al., 2018).

Edges in these contemporaneous networks represent effects that happen at a faster time-scale than the lag between the measurements. In my view, these contemporaneous networks are more likely to be valid estimations of attitude networks than the time-lagged networks, but one also has to estimate the contemporaneous attitude networks with caution, because also here limited variance causes problems.

The second relatively popular methodology to estimate individual networks is based on perceived causal relations (PCR; Frewen, Allen, Lanius, & Neufeld, 2012). The PCR methodology directly asks individuals about their network connections and has been applied to estimate networks of psychopathological symptoms and it would in principle also be possible to apply this method to attitude networks. However, PCR relies on a strong assumption – the assumption that individuals have knowledge of their own network structure. In the case of psychopathological symptoms this assumption might be reasonable to make (e.g., I can probably relatively accurately report how tired a sleepless night makes me), but for connections between attitude elements such an assumption might not pan out (e.g., it might be difficult to report how much judging a person as caring influences my judgment of this person being honest). Nonetheless, future research might do well in exploring under what circumstance (if any) PCR is useful in estimating individual attitude networks.

9.4 The Nature of Attitude Revisited

In the Introduction of this dissertation I discussed the question what an attitude is and argued that several prominent models on attitude rest on seemingly contradictory assumptions about the nature of attitude (e.g., whether attitudes are constructed on the spot or are stable representations in memory). Throughout this dissertation, I showed that the approach formulated here is able to unify several of these different assumptions. First, I have shown that while attitudes differ considerably in their strength (i.e., some attitudes are durable and impactful, while other attitudes are fluctuating and inconsequential) these differences arise from the same process – differences in connectivity and/or the dependence parameter of attitude networks (see Chapters 2, 4, 6). Second, low entropy attitudes *appear* to be stable representations in memory, because the strong dependency of attitude elements holds them in the same attitudinal macrostate (see Chapter 7). Third, what sometimes seems to be the result of two independent dual-processes (e.g., heuristic versus systematic processing of arguments) can be modeled based on the same process operating under different conditions (see Chapters 6 and 7).

Apart from integrating several different assumptions on attitude, the approach used in this dissertation also has implications for how we might view the nature of attitude. A central assumption of the formalisms developed in this dissertation is that studying the interactions between attitude elements provides us with more knowledge about attitudes than studying the underlying properties of attitude elements. If we take this assumption a step further we might even conclude that attitudes, or their underlying elements, have no nature in itself but that these interactions is all there is to know about attitude. From this perspective, specifying the nature of attitude is nothing more than a heuristic to summarize given properties of attitudes. For example, talking about attitudes as stable representations in memory provides an accurate description of the dynamic behavior and causal impact of strongly connected attitude networks. However, this does not imply that attitudes are in fact stable representations in memory. Similarly, it makes sense to talk of heuristic versus systematic processing of arguments as this provides an accurate description of the behavior of the attitude network under different circumstances. Again, however, this does not imply that there are actually two different processes at play.

The assumption that the nature of attitudes lies in their structure has several implications regarding fundamental questions on attitudes. In the following I discuss some of these implications. While I discuss these implications in the context of attitudes, it is my view that these implications can be also applied to psychological constructs more generally. I first discuss measurement from the perspective of this dissertation's approach. Second, I discuss how structural properties, such as attitudes, can have causal impact. Third, I place the current dissertation's approach into a broader context by linking it to recent developments in philosophy of science.

9.4.1 Measurement

As almost all psychological measurement, the measurement of attitude relies on the assumption that attitudes (and psychological constructs more generally) can be represented by a score that is independent of measurement (i.e., the construct score[1]) and that the construct score causally in-

[1] Note that I use the term construct score to refer to a score on a psychological construct (e.g., latent variable) instead of the more often used term true score. The reason for this is that the term construct score reflects a score on a property that is not solely a function of the test score and one that must have theoretical meaning. The term true score reflects the individual mean score on a given test if we were able to administer this test an infinitely

fluences the measurement outcome (Bollen, 1989; Bollen & Lennox, 1991; Borsboom et al., 2003; Nunnally, 1978). Both the central notions in psychological measurement of measurement error and measurement bias rely on the assumption of the construct score as measurement error is defined as random variation around the implied expected test score, given the construct score and the measurement model; measurement bias is defined as the systematic deviation from this implied expected test score. However, assuming that the formalisms developed in this dissertation are a somewhat valid description of reality, a construct score independent of measurement does not exist and the relation between psychological measurement and construct is not a causal, but a synergistic, one. Based on the AE framework, entropy of attitudes varies (partly) depending on the measurement instrument and because of that attitude scores cannot be interpreted as existing independent of measurement. While we formulated the AE framework in the context of attitudes, the generality of its principles is likely to extend to any form of mental representation (such as measurement of personality and individual difference measures more generally), implying that measurement influences the entropy of mental representations.

The implication that construct scores of mental representations do not exist independently of measurement does away with the classic definitions of measurement error and measurement bias. If a construct score does not exist, it cannot imply a definite expected test score, and as a result there can be no deviation from an expected test score. Logically, therefore, the AE framework implies that the very concepts of measurement error and measurement bias are inappropriate in the context of attitude measurement and that these concepts should be conceptualized in a new way. Regarding measurement error, the AE framework implies that random variation should be studied as a substantial property of constructs instead of defining it as mere noise, because treating random variation as something exogenous to the measured property results in loss of substantial information on the measured property. For example, as I showed in Chapter 5, higher levels of randomness in responses to measurements of attitudes toward presidential candidates are associated with low value of the attitude in predicting voting decisions. I therefore suggest that the common practice of treating random variation as measurement error is misguided and that such practice is only reasonable if one has a mechanistic theory

amount of times (Lord & Novick, 1968). True scores, however, are not interesting from a theoretical point of view, because any test – even tests that do not measure anything meaningful – have a true score (Borsboom & Mellenbergh, 2002).

on why random responses should be independent of the measured construct (see also van Geert & van Dijk, 2015 for a similar point regarding the measurement of developmental dynamics).

The above reasoning also leads to the conclusion that measurement error might be much less common than generally thought. I would argue that indeed true measurement error is a small fraction that can explain random variation and that all other random variation should be explicitly modeled. In addition to the measured construct showing inherent random variation, different measurement situations can create more or less fluctuating responses. For example, time pressure or fatigue might cause more random responses and it is my view that such external factors cannot be discarded as measurement error. Observing less consistent responses on attitude questionnaires when someone is tired or under time pressure might, for example, indicate that in such a situation the individual is more easily persuaded to adopt a different attitude (see also Chapter 7). Assuming that measurement error causes variation should thus be limited to such situations in which one can be reasonably certain that the variation is neither caused by the measured construct itself, the measurement context, or the measurement instrument. Limiting measurement error to such a narrow definition might even result in measurement error being so small that it becomes negligible and that assuming no measurement error might be a reasonable approximation in many circumstances – however, ultimately the question how pronounced measurement error is might be answered empirically by systematically controlling for all factors that might cause random variation.

Regarding measurement bias, the AE framework implies that the current definition of measurement bias rests on the untenable assumption that construct scores exist independent of measurement and the question arises whether it is possible to define measurement bias without relying on construct scores. My answer to this question is a simple "no" – strictly speaking measurement bias does not exist in the measurement of attitudes (and related constructs). In my view psychological studies assessing mental representations can be seen as a simulation of a real-world analogue (see also Brunswik, 1955; Lazer & Friedman, 2007) and the relevant question becomes to which extent any given study represents a valid simulation of the researched question. Regarding measurement, the question therefore becomes whether real-world analogues of the measurement process exist. In many cases, one is probably able to make the argument that measurement is analogous to a conversation an individual might have about the content of the measured property (e.g., when an individual answers the extraversion item whether one likes parties, probably similar processes are set in motion as when one would for some reason have a conversation about parties) or to a lesser extent to spontaneously think about the content of the measured property. However, when one wants to make the case that a given measurement has such a real-world analogue one has to specify whether administering the questionnaire is likely to cover a topic individuals talk or think regularly about or whether it is likely that an individual has never talked or thought about the topic, because the processes set in motion by the measurement will likely differ markedly between the two situations. In the situation when the item asks something the individual has talked or thought about frequently it is likely that the response is based on these earlier conversations, while in the situation when the item asks something the individual has never thought about, the answer to the item must be constructed on the spot.

I would suggest that measurement bias can be reconceptualized as inference bias, because it can represent processes one does not want to simulate in the current study. However, processes that lead to inference bias in one study may not lead to inference bias in another study. Take for example the tendency of individuals to respond in a socially desirable way (e.g., Crowne & Marlowe, 1964) in the context of measuring expressions of prejudice. Whether one wants to reduce the effect of social desirability depends on which form of prejudice expression one wants to simulate. While in some situations it is necessary to reduce the motivation to respond in a socially desirable way (e.g., when one wants to simulate the influence of prejudice on voting decisions), in other situations social desirability might be a relevant factor in simulating the studied situation (e.g., when one wants to simulate under which circumstances individuals are willing to express stereotypes). This example illustrates that while in some studies controlling for a given so-called measurement bias increases the validity of the given study, in other studies controlling for a given measurement bias can *reduce* validity, because one excludes a relevant process of the simulated situation.

9.4.2 Mental Causation

Treating attitudes as structures in which the constituting elements have no nature in itself creates a puzzle regarding how attitudes can have causal power – how can emergent higher-level structures influence lower level physical properties? This question mirrors the general puzzle of mental causation, which is a special form of downward causation (i.e., how do higher level properties such as mental states have causal influence on lower-level physical states?; Campbell, 1974). A possible answer how attitudes (and similar psychological constructs) can have causal power might lie in a recently proposed account of downward causation in biological systems, which holds that higher levels of organizations coarse-grain their macroscopic features (Flack, 2017). Coarse-graining refers to a commonly used technique in physics to derive an effective theory of a system (i.e., a theory that lets us predict the behavior of a system without knowing the system's exact microscopic configuration) by smoothing-out irrelevant noise of the macroscopic details of the system (e.g., to derive a system's temperature we only need to know the average speed of the system's particles and can ignore fluctuations around this average). Such a representation is often a more effective way to predict the system's behavior than relying on the fine-grained representation of the system (e.g., measuring temperature is an easier way to predict a system's behavior than by measuring each particle's speed). The reason why coarse-graining might be able to explain downward causation in biological systems is that it seems likely that biological systems use coarse-graining of their environment to make predictions and then adapt to these predictions. An example of this process is how macaques evaluate the power structure of their social system. Macaques estimate their standing in their group by coarse-graining how many fights they win and how many other macaques in their group show submissive behavior towards them (Flack, 2012, 2017). Coarse-graining these observations is advantageous, because it results in a simpler representation in memory and because it smoothes-out unimportant fluctuations (e.g., a fight outcome that is unrepresentative). Downward causation might result from these processes, because the emergent stable network of power estimations in the macaque group results in the macaques confirming to this emergent structure (e.g., two macaques that differ substantially in power are unlikely to engage in a fight).

Coarse-graining has a direct connection to entropy-reduction, because smoothing-out fluctuations leads to a representation lower in entropy than a representation that takes into account all fluctuations. Based on this we can reformulate the AE framework's principle that attention and thought lead to lower entropy representations to the principle that attention and thought lead to more coarse-grained representations. Therefore, attention and thought create the possibility of mental causation, because they force the mental system to filter out irrelevant variations and to create a simple coarse-grained representation to which the individual then adapts. For example, coarse-graining an attitude representation aids the decision process, because it enables the individual to arrive at a representation that unambiguously favors one option over the other. This reasoning fits findings on constraint-satisfaction processes prior to decision-making, which show that individuals interpret conflicting information in accordance with the overall evaluation they lean towards (e.g., Simon, Snow, & Read, 2004; Simon & Spiller, 2016). Also, the finding that mere measurement of intentions results in higher intention-behavior consistency (e.g., Godin, Sheeran, Conner, & Germain, 2008; Morwitz & Fitzsimons, 2004) can be explained in terms of the coarse-graining account – measurement leads to more coarse-grained representations, which in turn are more likely to have causal effects on subsequent behaviors.

9.4.3 Structural Realism

Taking a structural approach to theorizing on attitudes and related concepts echoes relatively recent advances in philosophy of science on structural realism. Structural realism is based on the idea that successful scientific theories might not *directly* describe the world correctly by illuminating the inherent nature of theoretical constructs such as particles, genes, or attitudes (as is assumed by entity realism). Established scientific theories, however, describe structural relations between constructs accurately – the structural properties described by such theories are objectively real (Ladyman, 1998). Accordingly, I would argue the only thing we can ever expect from a theory of attitudes is that it accurately describes the structural properties of attitude and that the nature of the constituents making up this structure will always remain elusive.

Structural realism is also in line with one of the most popular approaches to philosophy of mind – Daniel Dennett's (1987, 1991) theory of the intentional stance. The theory of the intentional stance assumes that to predict individual's behavior one can best do so by focusing on the beliefs and intentions of the individual. These beliefs and desires are considered to be real exactly because they have explanatory power (e.g., explaining why someone carries out a certain behavior is most straightforwardly done by analyzing possible beliefs and intentions toward this behavior) but they are not necessarily reducible to a biological basis (e.g., specific state of the brain). The theory of the intentional stance thus agrees well with structural realism, because it relies heavily on judging beliefs and intentions as real on structural aspects and at least entertains the possibility that the realness of a psychological state does not depend on entity realism.

9.5 Conclusion

Is psychology simple or complex? While most scholars would confidently state that psychology is extremely complex, I have shown in this dissertation that at least some of the complexity of one of psychology's core constructs can be explained by relying on simple principles. While I applied these principles to attitude, the generality of the principles makes it also likely that they extent to other domains and might serve as general blueprints of the dynamics of mental representations. There is no a priori reason to assume that only attitude elements interact with each other and other mental representations do not (in fact there is some evidence to the contrary; see Cramer et al., 2012) and it seems also unlikely that such interactions only increase for attitude elements when attention and thought are directed at them and not for other mental representations. I look forward to further explorations of the formalisms developed in this dissertation and I am optimistic that such explorations will not only advance the theoretical understanding of attitude but also the theoretical understanding of the human mind in general.

Part V

Appendices

References

Abelson, R. P. (1995). Attitude extremity. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 25–41). Hillsdale, NJ: Lawrence Erlbaum.

Abramowitz, A. I. (1995). It's abortion, stupid: Policy voting in the 1992 presidential election. *Journal of Politics*, *57*, 176–186.

Achard, S., Salvador, R., Whitcher, B., Suckling, J., & Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *The Journal of Neuroscience*, *26*, 63–72.

Achen, C. (1975). Mass political attitudes and the survey response. *American Political Science Review*, *69*, 1218–1231.

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for boltzmann machines. *Cognitive Science*, *9*, 147–169.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211.

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*, 888–918.

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Review of Modern Physics*, *74*, 47–97.

Aldrich, J. H., Sullivan, J. L., & Borgida, E. (1989). Foreign affairs and issue voting: Do presidential candidates "waltz before a blind audience?". *American Political Science Review*, *83*, 123–141.

Alligood, K. T., Sauer, T. D., Yorke, J. A., & Crawford, J. D. (1997). *Chaos: An introduction to dynamical systems*. New York, NY: Springer.

Alwin, D. F., & Krosnick, J. A. (1991). Aging, cohorts, and the stability of sociopolitical orientations over the life span. *American Journal of Sociology*, *97*, 169–195.

Amestoy, P. R., Azzalini, A., Badics, T., Benison, G., Böhm, W., Briggs, K., ... Yang, C. (2018). *igraph: Network analysis and visualization (R package version 1.2.1) [computer software]*. Retrieved from https://cran.r-project.org/web/packages/qgraph/index.html

Anderson, N. H. (1971). Integration theory and attitude change. *Psychological Review*, *78*, 171–206.

Anderson, P. W. (1972). More is different. *Science*, *177*, 393–396.

Ansolabehere, S., Rodden, J., & Snyder, J. M. (2008). The strength of issues: Using multiple measures to gauge preference stability, ideological constraint, and issue voting. *American Political Science Review*, *102*, 215–232.

Antonio, A. L., Chang, M. J., Hakuta, K., Kenny, D. A., Levin, S., & Milem, J. F. (2004). Effects of racial diversity on complex thinking in college students. *Psychological Science*, *15*, 507–510.

Armitage, C. J., & Arden, M. A. (2007). Felt and potential ambivalence across the stages of change. *Journal of Health Psychology*, *12*, 149–158.

Armitage, C. J., & Conner, M. (2000). Attitudinal ambivalence: A test of three key hypotheses. *Personality and Social Psychology Bulletin*, *26*, 1421–1432.

Baas, M., & De Dreu, B. A., C. K. W .and Nijstad. (2008). A meta-analysis of 25 years of mood-creativity research: Hedonic tone, activation, or regulatory focus? *Psychological Bulletin*, *134*, 779–806.

Bagozzi, R. P. (1981). Attitudes, intentions, and behavior: A test of some key hypotheses. *Journal of Personality and Social Psychology*, *41*, 607–627.

Bagozzi, R. P., & Baumgartner, J. (1986). The level of effort required for behaviour as a moderator of the attitude-behaviour relation. *European Journal of Social Psychology*, *20*, 45–59.

Bagozzi, R. P., & Burnkrant, R. E. (1979). Attitude organization and the attitude-behavior relationship. *Journal of Personality and Social Psychology*, *37*, 913–929.

Bagozzi, R. P., Tybout, A. M., Craig, C. S., & Sternthal, B. (1979). The construct validity of the tripartite classification of attitudes. *Journal of Marketing Research*, *16*, 88–95.

Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W. Y., Franklin, M. S., & Schooler, J. W. (2012). Inspired by distraction: Mind wandering facilitates creative incubation. *Psychological Science*, *23*, 1117–1122.

Barabási, A.-L. (2011). The network takeover. *Nature Physics*, *8*, 14–16.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–512.

Barabási, A.-L., & Zoltán, N. O. (2004). Network biology: Understanding the cell's functional organization. *Nature Review Genetics*, *5*, 101–113.

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*, *46*, 668–688.

Barden, J., Rucker, D. D., & Petty, R. E. (2005). "saying one thing and doing another": Examining the impact of event order on hypocrisy judgments of others. *Personality and Social Psychology Bulletin*, *31*, 1463–1474.

Barden, J., Rucker, D. D., Petty, R. E., & Rios, K. (2014). Order of actions mitigates hypocrisy judgments for ingroup more than outgroup members. *Group Processes & Intergroup Relations*, *17*, 590–601.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology*, *62*, 893–912.

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*, 53–58.

Barrat, A., Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 3747–3752.

Bartels, L. M. (2000). Partisanship and voting behavior, 1952-1996. *American Journal of Political Science*, *44*, 35–50.

Bartels, M., Rietveld, J. H., van Baal, G. C. M., & Boomsma, D. I. (2002). Genetic and environmental influences on the development of intelligence. *Behavior Genetics*, *32*, 236–249.

Bassili, J. N. (1996). Meta-judgmental versus operative indexes of psychological attributes: The case of measures of attitude strength. *Journal of Personality and Social Psychology*, *71*, 637–653.

Beckers, T., De Vicq, P., & Baeyens, F. (2009). Evaluative conditioning is insensitive to blocking. *Psychologica Belgica*, *49*, 41–57.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, *36*, 192–236.

Bizer, G. Y., Krosnick, J. A., Holbrrok, A. L., Wheeler, S. C., Rucker, D. D., & Petty, R. (2004). The impact of personality on cognitive, behavioral, and affective political processes: The effects of need to evaluate. *Journal of Personality*, *72*, 995–1027.

Bizer, G. Y., Larsen, J. T., & Petty, R. (2011). Exploring the valence-framing effect: Negative framing enhances attitude strength. *Political Psychology*, *32*, 59–80.

Bizer, G. Y., & Petty, R. (2005). How we conceptualize our attitudes matters: The effects of valence framing on the resistance of political attitudes. *Political Psychology*, *26*, 553–568.

Block, N. (2005). Two neural correlates of consciousness. *Trends in Cognitive Sciences*, *9*, 46–52.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, *30*, 481–499.

Bollen, K. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*, 305–314.

Boltzmann, L. (1877). Über die Beziehung zwischen dem zweiten Hauptsatz der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen ber das Wärmegleichgewicht [On the relationship between the second law of the mechanical theory of heat and the probability calculus]. *Wiener Berichte*, *76*, 373–435.

Boninger, D. S., Krosnick, J. A., Berent, M. K., & Fabrigar, L. R. (1995). The causes and consequences of attitude importance. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 159–189). Hillsdale, NJ: Lawrence Erlbaum.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, *323*, 892–895.

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, *64*, 1089–1108.

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91-121.

Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PLoS ONE*, *6*, e27407.

Borsboom, D., & Mellenbergh, G. J. (2002). True scores, latent variables, and constructs: A comment on schmidt and hunter. *Intelligence*, *30*, 505–514.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*, 1061–1071.

Borsboom, D., Rhemtulla, M., Cramer, A., Van der Maas, H., Scheffer, M.,

& Dolan, C. (2016). Kinds versus continua: a review of psychometric approaches to uncover the structure of psychiatric constructs. *Psychological Medicine*, *46*, 1567–1579.

Boschloo, L., van Borkulo, C. D., Borsboom, D., & Schoevers, R. A. (2016). A prospective study on how symptoms in a network predict the onset of depression. *Psychotherapy and Psychosometics*, *85*, 183–184.

Brandes, U. (2001). A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology*, *25*, 163–177.

Brandt, M. J., Sibley, C. G., & Osborne, D. (2019). What is central to belief system networks? *Personality and Social Psychology Bulletin*. (Advance online publication)

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology*, *47*, 1191–1205.

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., ... Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS ONE*, *8*, e60188.

Brown, L. (1986). *Fundamentals of statistical exponential families*. Hayward, CA:Institute of Mathematical Statistics.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193.

Burcusa, S. L., & Iacono, W. G. (2007). Risk for recurrence in depression. *Clinical Psychology Review*, *27*, 959–985.

Busemeyer, J. R., & Wang, Z. (2018). Hilbert space multidimensional theory. *Psychological Review*, *125*, 572–591.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, *115*, 401–423.

Cacioppo, J. T., Petty, R. E., & Green, T. R. (1989). From the tripartite model to the homeostasis model of attitudes. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 275–310). Hillsdale, NJ: Lawrence Erlbaum.

Campbell, D. T. (1974). 'downward causation' in hierarchically organised biological systems. In *Studies in the philosophy of biology* (pp. 179–186). London: Palgrave.

Cannell, C., Miller, P., & Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 389–437). San Francisco: Jossey-Bass.

Cantril, H. (1946). The intensity of an attitude. *The Journal of Abnormal and Social Psychology*, *41*, 129–135.

Carlson, T. N., & Settle, J. E. (2016). Political chameleons: An exploration of conformity in political discussions. *Political Behavior*, *38*, 817–859.

Carmines, E., & Stimson, J. (1980). The two faces of issue voting. *American Political Science Review*, *74*, 78–91.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*, 747–749.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). New York, NY: Guilford Press.

Chaiken, S., Pomerantz, E. M., & Giner-Sorolla, R. (1995). Structural consistency and attitude strength. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 387–412). Hillsdale, NJ: Lawrence Erlbaum.

Chang, W. (2018). *shiny: Web application framework for R (R package version (1.1.0) [computer software].* Retrieved from `https://cran.r-project.org/web/packages/shiny/index.html`

Chen, J., & Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*, 759–771.

Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality and Social Psychology*, *69*, 318–328.

Clark, J. K., Wegener, D. T., & Fabrigar, L. R. (2008). Attitudinal ambivalence and message-based persuasion: Motivated processing of proattitudinal information and avoidance of counterattitudinal information. *Personality and Social Psychology Bulletin*, *34*, 565–577.

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, *70*, 066111.

Comte, A. (1842). *The positive philosophy*. New York: Blanchard.

Conner, M., Godin, G., Sheeran, P., & Germain, M. (2013). Some feelings are more important: Cognitive attitudes, affective attitudes, anticipated affect, and blood donation. *Health Psychology*, *32*, 264–272.

Conner, M., & Sparks, P. (2002). Ambivalence and attitudes. *European*

Review of Social Psychology, *12*, 37–70.

Converse, P. E. (1970). Attitudes and non-attitudes: Continuation of a dialogue. In E. R. Tufte (Ed.), *The quantative analysis of social problems* (pp. 168–189). Reading, MA: Addison-Wesley.

Corneille, O., & Stahl, C. (2018). Associative attitude learning: A closer look at evidence and how it relates to attitude models. *Personality and Social Psychology Review*. (Advance online publication)

Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, A. O. J. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality*, *54*, 13–29.

Cramer, A. O. J., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PLoS ONE*, *11*, e0167490.

Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., … Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, *26*, 414–431.

Cramer, A. O. J., Waldorp, L. J., van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, *33*, 137–193.

Crano, W. D. (1997). Vested interest, symbolic politics, and attitude-behavior consistency. *Journal of Personality and Social Psychology*, *72*, 485–491.

Crites, S. L., Fabrigar, L. R., & Petty, R. E. (1994). Measuring the affective and cognitive properties of attitudes: Conceptual and methodological issues. *Personality and Social Psychology Bulletin*, *20*, 619–634.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence*. New York, NY: Wiley.

Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, *1695*.

Cunningham, W. A., & Luttrell, A. (2015). Attitudes. In A. W. Toga (Ed.), *Brain mapping: An encyclopedic reference. vol. 3* (pp. 235–239). Amsterdam: Elsevier.

Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: A social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *11*, 97–104.

Cunnningham, W. A., Dunfield, K. A., & Stillman, P. E. (2013). Emotional states from affective dynamics. *Emotion Review*, *5*, 344–355.

Cunnningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *Journal of Cognitive Neuroscience*, *16*, 1717–1729.

Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. J. (2016). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychological Review*, *123*, 2–22.

Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2017). Network analysis on attitudes: A brief tutorial. *Social Psychological and Personality Science*, *8*, 528–537.

Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2018a). The attitudinal entropy (ae) framework as a general theory of individual attitudes. *Psychological Inquiry*, *29*, 175-193.

Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2018b). A network perspective on attitude strength: Testing the connectivity hypothesis. *Social Psychological and Personality Science*. (Advance online publication)

Dalege, J., Borsboom, D., van Harreveld, F., Waldorp, L. J., & van der Maas, H. L. J. (2017). Network structure explains the impact of attitudes on decisions. *Scientific Reports*, *7*, 4909.

Daley, M. (2018). Circe's victims: Are we too easily seduced by the siren song of mathematical physics? *Psychological Inquiry*, *29*, 194-195.

De Boeck, P., Wilson, M., & Acton, G. S. (2005). A conceptual and psychometric framework for distinguishing categories and dimensions. *Psychological Review*, *112*, 129–158.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, *135*, 347–368.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and disklikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*, 853–869.

Dennett, D. C. (1987). *The intentional stance*. Cambridge: MIT Press.

Dennett, D. C. (1991). Real patterns. *The Journal of Philosophy*, *88*, 27–51.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*, 269–271.

Dunne, J. A., Williams, R. J., & Martinez, N. D. (2002). Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 12917–12922.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Orlando, FL: Harcourt Brace Jovanovich.

Eagly, A. H., & Chaiken, S. (1995). Structural consitency and attitude strength. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 413–432). Hillsdale, NJ: Lawrence Erlbaum.

Eagly, A. H., & Chaiken, S. (2007). The advantages of an inclusive definition of attitude. *Social Cognition*, *25*, 582–602.

Eaton, A. A., Visser, P. S., Krosnick, J. A., & Anand, S. (2009). Social power and attitude strength over the life course. *Personality and Social Psychology Bulletin*, *35*, 1646–1660.

Eaves, L. J., Eysenck, H. J., & Martin, N. G. (1989). *Genes, culture and personality: An empirical approach*. San Diego, CA: Academic Press.

Edwards, K. (1990). The interplay of affect and cognition in attitude formation and change. *Journal of Personality and Social Psychology*, *59*, 202–216.

Ehret, P. J., Monroe, B. M., & Read, S. J. (2015). Modeling the dynamics of evaluation: A multilevel neural network implementation of the iterative reprocessing model. *Personality and Social Psychology Review*, *19*, 148–176.

Epskamp, S. (2015). *Isingsampler: Sampling methods and distribution functions for the ising model (R package version 0.2) [computer software].* Retrieved from `https://cran.r-project.org/web/packages/IsingSampler/index.html`

Epskamp, S. (2016). *Regularized gaussian psychological networks: Brief report on the performance of extended bic model selection.* Retrieved from `http://arxiv.org/abs/1606.05771`

Epskamp, S. (2017). *bootnet: Bootstrap methods for various network estimation routines (R package version 1.0.1) [computer software].* Retrieved from `https://cran.r-project.org/web/packages/bootnet/index.html`

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*, 195–212.

Epskamp, S., Costantini, G., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & D., B. (2018). *qgraph: Graph plotting methods, psychometric data visualization and graphical model estimation (R package version 1.5) [computer software].* Retrieved from `https://cran.r-project.org/web/packages/qgraph/index.html`

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*, 1–18.

Epskamp, S., Maris, G., Waldorp, L., & Borsboom, D. (2018). Network psychometrics. In P. Irwing, D. Hughes, & T. Booth (Eds.), *The wiley handbook of psychometric testing, 2 volume set: A multidisciplinary reference on survey, scale and test development.* New York, NY: Wiley.

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, 1–28.

Erdos, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae*, *6*, 290–297.

Erikson, R. S., & Wlezien, C. (1999). Presidential polls as a time series: The case of 1996. *Public Opinion Quarterly*, *63*, 163–177.

Eschleman, K. J., Bowling, N. A., & Judge, T. A. (2015). The dispositional basis of attitudes: A replication and extension of hepler and albarracín (2013). *Journal of Personality and Social Psychology*, *108*, e1–e15.

Fabrigar, L. R., MacDonald, T. K., & Wegener, D. T. (2005). The structure of attitudes. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 79–125). Mahwah, NJ: Erlbaum.

Fabrigar, L. R., & Petty, R. E. (1999). The role of the affective and cognitive bases of attitudes in susceptibility to affectively and cognitively based persuasion. *Personality and Social Psychology Bulletin*, *35*, 363–381.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The mode model as an integrative framework. *Advances in Experimental Social Psychology*, *23*, 75–109.

Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247–282). Hillsdale, NJ: Lawrence Erlbaum.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, *25*, 603–637.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013–1027.

Fazio, R. H., & Olson, M. A. (2003a). Attitudes: Foundations, functions, and consequences. In M. A. Hogg & J. Cooper (Eds.), *The handbook of social psychology* (pp. 163–190). London: Sage.

Fazio, R. H., & Olson, M. A. (2003b). Implicit measures in social cognition.

research: Their meaning and use. *Annual Review of Psychology*, *54*, 297–327.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *51*, 505–514.

Fazio, R. H., & Williams, C. J. (1986). Attitude accessibility as a moderator of the attitude-perception and attitude-behavior relations: An investigation of the 1984 presidential election. *Journal of Personality and Social Psychology*, *51*, 505–514.

Fazio, R. H., & Zanna, M. (1981). Direct experience and attitude-behavior consistency. *Advances in Experimental Social Psychology*, *14*, 161–202.

Ferrer i Cancho, R., & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*, 2261–2265.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, *58*, 203–210.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Fiske, S. T. (2004). Mind the gap: In praise of informal sources of formal theory. *Personality and Social Psychology Review*, *8*, 132–137.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*, 878–902.

Flack, J. C. (2012). Multiple time-scales and the developmental dynamics of social systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 1802–1810.

Flack, J. C. (2017). Coarse-graining as a downward causation mechanism. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *375*, 20160338.

Flay, B. R. (1978). Catastrophe theory in social psychology: Some applications to attitudes and social behavior. *Behavioral Science*, *23*, 335–350.

Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. In J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, & A. Culotta (Eds.), *Advances in neural information processing systems 23* (pp. 604–612). Red Hook, NY:

Curran Associates, Inc.
Fraiman, D., Balenzuela, P., Foss, J., & Chialvo, D. R. (2009). Ising-like dynamics in large-scale functional brain networks. *Physical Review E*, *79*, 061922.
Fredrickson, G. H., & Andersen, H. C. (1984). Kinetic Ising model of the glass transition. *Physical Review Letters*, *53*, 1244–1247.
Freeman, L. C. (1978). Centrality in social networks: Conceptual clarification. *Social Networks*, *1*, 215–239.
Frewen, P. A., Allen, S. L., Lanius, R. A., & Neufeld, R. W. J. (2012). Perceived causal relations: novel methodology for assessing client attributions about causal associations between variables including symptoms and functional impairment. *Assessment*, *19*, 480–493.
Frewen, P. A., Schmittmann, V. D., Bringmann, L. F., & Borsboom, D. (2013). Perceived causal relations between anxiety, posttraumatic stress and depression: extension to moderation, mediation, and network analysis. *European Journal of Psychotraumatology*, *4*, 20656.
Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders*, *189*, 314–320.
Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*, 432–441.
Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*, 293–301.
Friston, K., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics*, *102*, 227–260.
Fruchterman, T., & Reingold, E. (1991). Graph drawing by force directed placement. *Software, Practice & Experience*, *21*, 1129–1164.
Fulker, D. W., DeFries, J. C., & Plomin, R. (1988). Genetic influence on general mental ability increases between infancy and middle childhood. *Nature*, *336*, 767–769.
Galam, S., Gefen, Y., & Shapir, Y. (1982). Sociophysics: A new approach of sociological collective behaviour. I. mean–behaviour description of a strike. *Journal of Mathematical Sociology*, *9*, 1–13.
Galdi, S., Arcuri, L., & Gawronski, B. (2008). Automatic mental associations predict future choices of undecided decision-makers. *Science*, *321*, 1100–1102.
Galesic, M., Olsson, H., & Rieskamp, J. (2018). A sampling model of social judgment. *Psychological Review*, *125*, 363–390.

Galesic, M., & Stein, D. (2019). Statistical physics models of belief dynamics: Theory and empirical tests. *Physica A: Statistical Mechanics and its Applications*, 275–294.

Gates, K. M., Henry, T., Steinley, D., & Fair, D. A. (2016). A monte carlo evaluation of weighted community detection algorithms. *Frontiers in neuroinformatics*, *10*, 45.

Gawronski, B. (2012). Back to the future of dissonance theory: Cognitive consistency as a core motive. *Social Cognition*, *30*, 652–668.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731.

Gawronski, B., & Bodenhausen, G. V. (2007). Unraveling the processes underlying evaluation: Attitudes from the perspective of the ape model. *Social Cognition*, *25*, 687–717.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology*, *44*, 59–127.

Gawronski, B., LeBel, E. P., & Peters, K. P. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, *2*, 181–193.

Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin*, *43*, 300–312.

Gawronski, B., & Strack, F. (2012). Cognitive consistency as a basic principle of social information processing. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 1–16). New York, NY: Guilford Press.

Gerring, J. (1997). Ideology: A definitional analysis. *Political Research Quarterly*, *50*, 957–994.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.

Gilmore, R. (1981). *Catastrophe theory for scientists and engineers*. New York: John Wiley.

Girvan, M., Newman, M. E. J., & Moore, C. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of the United States of America*, *99*, 7821–7826.

Glasman, L. R., & Albarracín, D. (2006). Forming attitudes that predict future behavior: A meta-analysis of the attitude-behavior relation. *Psychological Bulletin*, *132*, 778–822.

Glauber, R. J. (1963). Time–dependent statistics of the Ising model. *Journal of Mathematical Physics*, *4*, 294–307.
Godin, G., Sheeran, P., Conner, M., & Germain, M. (2008). Asking questions changes behavior: Mere measurement effects on frequency of blood donation. *Health Psychology*, *27*, 179-184.
Goel, S., Mason, W., & Watts, D. J. (2010). Real and perceived attitude agreement in social networks. *Journal of Personality and Social Psychology*, *99*, 611-621.
Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS ONE*, *12*, e0174035.
Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.
Gross, S. R., Holtz, R., & Miller, N. (1995). Attitude certainty. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 215–245). Hillsdale, NJ: Lawrence Erlbaum.
Haddock, G., Zanna, M. P., & Esses, V. M. (1971). Validation of feeling, belief, and intention to act as three components of attitude and their contribution to prediction of contraceptive behavior. *Journal of Personality and Social Psychology*, *19*, 321–333.
Haddock, G., Zanna, M. P., & Esses, V. M. (1993). Assessing the structure of prejudicial attitudes: The case of attitudes toward homosexuals. *Journal of Personality and Social Psychology*, *65*, 1105–1118.
Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Methods*, *10*, 371-388.
Hariri, A. R., Mattay, V. S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., ... Weinberger, D. R. (2002). Serotonin transporter genetic variation and the response of the human amygdala. *Science*, *297*, 400–403.
Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, *135*, 555–588.
Haslbeck, J. (2017). *mgm: Estimating time-varying k-order mixed graphical models (R package version 1.2-2) [computer software].* Retrieved from `https://cran.r-project.org/web/packages/mgm/index.html`
Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. New York, NY: Springer.
Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with*

sparsity: The lasso and generalizations*. New York, NY: CRC Press.
Hebb, D. O. (1949). *The organization of behavior*. New York, NY: Wiley & Sons.
Hehman, E., Flake, J. K., & Calanchini, J. (2018). Disproportionate use of lethal force in policing is associated with regional racial biases of residents. *Social Psychological and Personality Science*, *9*, 393–401.
Heider, F. (1946). Attitudes and cognitive organization. *Advances in Experimental Social Psychology*, *21*, 107–112.
Heider, F. (1958). *The psychology of cognitive organization*. Hoboken, NJ: Wiley.
Heine, S. J., & Lehman, D. R. (1997). Culture, dissonance, and self-affirmation. *Personality and Social Psychology Bulletin*, *23*, 389–400.
Hepler, J., & Albarracín, D. (2013). Attitudes without objects: Evidence for a dispositional attitude, its measurement, and its consequences. *Journal of Personality and Social Psychology*, *104*, 1060–1076.
Herzog, H. A. (1993). "the movement is my life": The psychology of animal rights activism. *Journal of Social Issues*, *49*, 103–119.
Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, *52*, 1280-1300.
Higgins, E. T. (2004). Making a theory useful: Lessons handed down. *Personality and Social Psychology Review*, *8*, 138–145.
Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychological Review*, *119*, 304–320.
Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*, 1369–1385.
Holland, R. W., Verplanken, B., & van Knippenberg, A. (2002). On the nature of attitude-behavior relations: The strong guide, the weak follow. *European Journal of Social Psychology*, *32*, 869–876.
Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General*, *128*, 3–31.
Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*, 2554–2558.
Hopfield, J. J. (1984). Neurons with graded responses have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *81*,

3088–3092.

Hoshino-Browne, E., Zanna, A. S., Spencer, S. J., Zanna, M. P., Kitayama, S., & Lackenbauer, S. (2005). On the cultural guises of cognitive dissonance: The case of easterners and westerners. *Journal of Personality and Social Psychology*, *89*, 294–310.

Houston, D. A., & Fazio, R. H. (1989). Biased processing as a function of attitude accessibility: Making objective judgments subjectively. *Social Cognition*, *7*, 51–66.

Howe, L. C., & Krosnick, J. A. (2017). Attitude strength. *Annual Review of Psychology*, *68*, 327–351.

Humphries, M. D., & Gurney, K. (2008). Network "small-world-ness": A quantitative method for determining canonical network equivalence. *PLoS ONE*, *3*, e0002051.

Huskinson, T. L. H., & Haddock, G. (2004). Individual differences in attitude structure: Variance in the chronic reliance on affective and cognitive information. *Journal of Experimental Social Psychology*, *40*, 82–90.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York, NY: Basic Books.

Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus [Contribution to the theory of ferromagnetism]. *Zeitschrift für Physik*, *31*, 253–258.

Jacoby, W. G. (2010). Policy attitudes, ideology and voting behavior in the 2008 election. *Electoral Studies*, *29*, 557–568.

James, W. (1890). *The principles of psychology*. Cambridge: Harvard University Press.

Janis, I. L. (1982). *Psychological studies of policy decisions and fiascos*. Boston: Houghton Mifflin.

Jaynes, E. T. (1965). Gibbs vs Boltzmann entropies. *American Journal of Physics*, *33*, 391–398.

Jonas, K., Broemer, P., & Diehl, M. (2000). Attitudinal ambivalence. *European Review of Social Psychology*, *11*, 35–74.

Jonas, K., Diehl, M., & Bromer, P. (1997). Effects of attitudinal ambivalence on information processing and attitude-intention consistency. *Journal of Experimental Social Psychology*, *33*, 190–210.

Jones, C. R., Olson, M. A., & Fazio, R. H. (2010). Evaluative conditioning: The "how" question. *Advances in Experimental Social Psychology*, *43*, 205–255.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*, 109–133.

Judd, C., & Krosnick, J. (1989). The structural bases of consistency among political attitudes: Effects of expertise and attitude importance. In A. Pratkanis, S. Breckler, & A. Greenwald (Eds.), *Attitude structure and function* (pp. 99–128). Hillsdale, NJ: Lawrence Erlbaum.

Judd, C., Krosnick, J., & Milburn, M. (1980). The structure of attitude systems in the general public: Comparisons of a structural equation model. *American Sociological Review*, *45*, 627–643.

Judd, C., Krosnick, J., & Milburn, M. (1981). Political involvement and attitude structure in the general public. *American Sociological Review*, *46*, 660–669.

Katz, D. (1960). The functional approach to the study of attitudes. *Public opinion quarterly*, *24*, 163–204.

Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution*. New York: Oxford University Press.

Kim, B. J., Liu, J., Um, J., & Lee, S. I. (2005). Instability of defensive alliances in the predator-prey model on complex networks. *Physical Review E*, *72*, 041906.

Kindermann, R., & Snell, J. L. (1980). *Markov random fields and their applications*. Providence, RI: American Mathematics Society.

Kitayama, S., Snibbe, A. C., Markus, H. R., & Suzuki, T. (2004). Is there any ”free” choice? Self and dissonance in two cultures. *Psychological Science*, *15*, 527–533.

Koestler, A. (1964). *The act of creation*. New York, NY: Macmillan.

Kolaczyk, E. D. (2009). *Statistical analysis of network data: Methods and models*. New York, NY: Springer.

Kraus, S. J. (1986). Attitudes and the prediction of behavior: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin*, *51*, 505–514.

Krieglmeyer, R., & Sherman, J. W. (2012). Disentangling stereotype activation and stereotype application in the stereotype misperception task. *Journal of Personality and Social Psychology*, *103*, 205–224.

Krosnick, J. A. (1988). The role of attitude importance in social evaluation: a study of policy preferences, presidential candidate evaluations, and voting behavior. *Journal of Personality and Social Psychology*, *55*, 196–210.

Krosnick, J. A. (1989). Attitude importance and attitude accessibility. *Personality and Social Psychology Bulletin*, *15*, 297–308.

Krosnick, J. A., & Alwin, D. F. (1989). Aging and susceptibility to attitude change. *Journal of Personality and Social Psychology*, *57*, 416–425.

Krosnick, J. A., Boninger, D. S., Chuang, Y. C., Berent, M. K., & Carnot, C. G. (1993). Attitude strength: One construct or many related constructs? *Journal of Personality and Social Psychology, 65*, 1132–1151.

Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 1–24). Hillsdale, NJ: Lawrence Erlbaum.

Kruglanski, A. W. (2001). That "vision thing": The state of theory in Social and Personality Psychology at the edge of the new millennium. *Journal of Personality and Social Psychology, 80*, 871–875.

Kruglanski, A. W., Katarzyna, J., Chernikova, M., Milyavsky, M., Babush, M., Baldner, C., & Pierro, A. (2015). The rocky road from attitudes to behaviors: Charting the goal systemic course of actions. *Psychological Review, 122*, 598–620.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review, 103*, 284–308.

Ladyman, J. (1998). What is structural realism? *Studies in History and Philosophy of Science, 29*, 409–424.

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7*, 12–18.

Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences, 7*, 494–501.

LaPiere, E. (1934). Attitudes vs. actions. *Social Forces, 13*, 230–237.

Larson, C. L., Schaefer, H. S., Siegle, G. J., Jackson, C. A. B., Anderle, M. J., & Davidson, R. J. (2006). Fear is fast in phobic individuals: Amygdala activation in response to fear-relevant stimuli. *Biological Psychiatry, 60*, 410–417.

Latané, B., & Nowak, A. (1994). Attitudes as catastrophes: From dimensions to categories with increasing involvement. In R. R. Vallaches & A. Nowak (Eds.), *Dynamical systems in social psychology* (pp. 219–249). San Diego, CA: Academic Press.

Laumann, E. O., Marsden, P. V., & Prensky, D. (1989). The boundary specification problem in network analysis. In L. C. Freeman, D. R. White, & A. K. Romney (Eds.), *Methods in social network analysis* (pp. 61–88). New Brusnwick, NJ: Transaction Publishers.

Lauritzen, S. L. (1996). *Graphical models*. Oxford, UK: Clarendon Press.

Lavine, H., Thomsen, C. J., Zanna, M. P., & Borgida, E. (1998). On the primacy of affect in the determination of attitudes and behavior: The moderating role of affective-cognitive ambivalence. *Journal of Experimental Social Psychology, 34*, 398–421.

Lazer, D., & Friedman, A. (2007). The network structure of exploration and exploitation. *Administrative Science Quarterly*, *52*, 667–694.

LeBel, E. P., & Paunonen, S. V. (2011). Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin*, *37*, 570–583.

Lewin, K. (1951). *Field theory in social science*. New York: Harper.

Little, R. J. A. (1988). Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics*, *6*, 287–296.

Liu, J. H., & Latané, B. (1998). The catastrophic linke between the importance and extremity of political attitudes. *Political Behavior*, *20*, 105–126.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Lusk, C. M., & Judd, C. M. (1988). Political expertise and the structural mediators of candidate evaluations. *Journal of Experimental Social Psychology*, *24*, 105–126.

Mackintosh, N. J. (1998). *Iq and human intelligence*. Oxford, England: Oxford University Press.

Manrubia, S. C., & Mikhailov, A. S. (1999). Mutual synchronization and clustering in randomly coupled chaotic dynamical networks. *Physical Review E*, *60*, 1579–1589.

March, D. S., Olson, M. A., & Gaertner, L. (2018). On logical and mathematical boxes: Does the attitudinal entropy framework expand our understanding of attitudes? *Psychological Inquiry*, *29*, 196-199.

Markus, G. B. (1982). Political attitudes during an election year: A report on the 1980 nes panel study. *American Political Science Review*, *76*, 538–560.

McGregor, I., Newby-Clark, I. R., & Zanna, M. P. (1999). ”Remembering” dissonance: Simultaneous accessibility of inconsistent cognitive elements moderates epistemic discomfort. In E. Harmon-Jones & J. Mills (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology* (pp. 325–352). Washington: American Psychological Association.

McGuire, W. J. (1990). Dynamic operations of thought systems. *American Psychologist*, *45*, 504–512.

McNally, R. J., Robinaugh, D. J., Wu, G. W. Y., Wang, L., Deserno, M. K., & Borsboom, D. (2015). Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical Psychological Science*, *3*, 836–849.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather:

Homophily in social networks. *Annual Review of Sociology*, *27*, 415–444.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir karl, sir ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806–834.

Millar, M. G., & Millar, K. U. (1996). Affective and cognitive responses and the attitude-behavior relation. *Journal of Experimental Social Psychology*, *32*, 561–579.

Millar, M. G., & Tesser, A. (1986). Thought-induced attitude change: The effects of schema structure and commitment. *Journal of Personality and Social Psychology*, *51*, 259–269.

Miller, W. E. (1991). Party identification, realignment , and party voting: Back to the basics. *American Political Science Review*, *85*, 557–568.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*, 183–198.

Mitchell, M. (2009). *Complexity: A guided tour*. Oxford: Oxford University Press.

Monroe, B. M. (2018). Progress in understanding requires modeling with faithfulness to the phenomenon, and precise mechanisms. *Psychological Inquiry*, *29*, 200-202.

Monroe, B. M., & Read, S. J. (2008). A general connectionist model of attitude structure and change: the ACS (Attitudes as Constraint Satisfaction) model. *Psychological Review*, *115*, 733–759.

Morris, J. S., Frith, C. D., Perrett, D. I., & Rowland, D. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, *383*, 812–815.

Morwitz, V. G., & Fitzsimons, G. J. (2004). The mere-measurement effect: Why does measuring intentions change actual behavior? *Journal of Consumer Psychology*, *14*, 64–74.

Murphy, K. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press.

Nadeau, R., & Lewis-Beck, M. S. (2001). National economic voting in U.S. presidential elections. *Journal of Politics*, *63*, 159–181.

Newby-Clark, I. R., McGregor, I., & Zanna, M. P. (2002). Thinking and caring about cognitive inconsistency: When and for whom does attitudinal ambivalence feel uncomfortable? *Journal of Personality and Social Psychology*, *82*, 157–166.

Newman, M. E. J. (2001). Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical Review E*, *64*,

16132.
Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, *69*, 066133.
Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of the United States of America*, *103*, 8577–8582.
Newman, M. E. J. (2010). *Networks: An introduction*. Oxford: Oxford University Press.
Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*, 026113.
Nohlen, H. U., van Harreveld, F., Rotteveel, M., Lelieveld, G. J., & Crone, E. A. (2013). Evaluating ambivalence: social-cognitive and affective brain regions associated with ambivalent decision-making. *Social Cognitive and Affective Neuroscience*, *9*, 924–931.
Nordgren, L. F., van Harreveld, F., & van der Pligt, J. (2006). Ambivalence, discomfort, and motivated information processing. *Journal of Experimental Social Psychology*, *42*, 252–258.
Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition*, *19*, 625–666.
Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences*, *15*, 152–159.
Nunnally, J. (1978). *Psychometric methods*. New York, NY: McGraw-Hill.
Olson, J. M., Vernon, P. A., Harris, J. A., & Jang, K. L. (2001). The heritability of attitudes: A study of twins. *Journal of Personality and Social Psychology*, *80*, 845–860.
Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*, 245–251.
Ostrom, T. M. (1969). The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology*, *5*, 12–30.
Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*, 171–192.
Oxley, D. R., Smith, K. B., Alford, J. R., Hibbing, M. V., Miller, J. L., Scalora, M., ... Hibbing, J. R. (2008). Political attitudes vary with physiological traits. *Science*, *321*, 1667–1670.
Palfrey, T. R., & Poole, K. T. (1978). Economic retrospective voting in

American national elections: A micro-analysis. *American Journal of Political Science*, *22*, 426–443.

Palfrey, T. R., & Poole, K. T. (1987). The relationship between information, ideology, and voting behavior. *American Journal of Political Science*, *31*, 511–530.

Papadopoulos, F., Kitsak, M., Serrano, M. A., Boguña, M., & Krioukov, D. (2012). Popularity versus similarity in growing networks. *Nature*, *489*, 537–540.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277–293.

Payne, B. K., Krosnick, J. A., Pasek, J., Lelkes, Y., Akhtar, O., & Tompson, T. (2010). Implicit and explicit prejudice in the 2008 american presidential election. *Journal of Experimental Social Psychology*, *46*, 367–374.

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*, 233–248.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 534–552.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York, NY: Cambridge University Press.

Petrocelli, J. V., Tormala, Z. L., & Rucker, D. D. (2007). Unpacking attitude certainty: attitude clarity and attitude correctness. *Journal of Personality and Social Psychology*, *92*, 30–41.

Petty, R. E., & Briñol, P. (2018). The aef: Reinforcing our knowledge about attitudes using a physics metaphor. *Psychological Inquiry*, *29*, 203-207.

Petty, R. E., & Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, *46*, 69–81.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*, 123–205.

Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, *41*, 847–855.

Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Consumer Research*, *10*, 135–146.

Petty, R. E., Haugtvedt, C. P., & Smith, S. M. (1995). Elaboration as a determinant of attitude strength: Creating attitudes that are persistent, resistant, and predictive of behavior. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 93–130). Hillsdale, NJ: Lawrence Erlbaum.

Phelps, E. A. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology*, *57*, 27–53.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. New York, NY: Basic Books.

Plotnitsky, A. (2012). *Niels bohr and complementarity: An introduction*. Berlin: Springer.

Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks. In P. Yolum, T. Güngör, F. Gürgen, & C. Özturan (Eds.), *International symposium on computer and information sciences* (pp. 284–293). Berlin: Springer.

Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, *36*, 255–274.

Priester, J. R., & Petty, R. E. (1996). The gradual threshold model of ambivalence: Relating the positive and negative bases of attitudes to subjective ambivalence. *Journal of Personality and Social Psychology*, *71*, 431–449.

Priester, J. R., & Petty, R. E. (2001). Extending the bases of subjective attitudinal ambivalence: Interpersonal and intrapersonal antecedents of evaluative tension. *Journal of Personality and Social Psychology*, *80*, 19–34.

Prislin, R. (1996). Attitude stability and attitude strength: One is enough to make it stable. *European Journal of Social Psychology*, *26*, 447–477.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rabinowitz, G., & MacDonald, S. E. (1989). A directional theory of issue voting. *American Political Science Review*, *83*, 93–121.

Read, S. J., Vanman, E. J., & Miller, L. C. (1997). Connectionism, parallel constraint satisfaction processes, and gestalt principles: (re) introducing cognitive dynamics to social psychology. *Personality and Social Psychology Review*, *1*, 26–53.

Refling, E. J., Calnan, C. M., Fabrigar, L. R., MacDonald, T. K., Johnson,

V. C., & Smith, S. M. (2013). To partition or not to partition evaluative judgments: Comparing measures of structural ambivalence. *Social Psychological and Personality Science*, *4*, 387–394.

Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reis & C. Judd (Eds.), *Handbook of research methods in social psychology* (pp. 190–222). New York, NY: Cambridge University Press.

Reynolds, C. (1987). Flocks, herds and schools: A distributed behavioral model. In *Siggraph '87: Proceedings of the 14th annual conference on computer graphics and interactive techniques* (pp. 25–34). New York, NY: Association for Computing Machinery.

Rios, K., & Roth, Z. C. (2018). Extending the attitudinal entropy framework to interpersonal phenomena: A commentary on dalege et al. (2018). *Psychological Inquiry*, *29*, 208-212.

Rohr, M., Degner, J., & Wentura, D. (2014). The "emotion misattribution" procedure: Processing beyond good and bad under masked and unmasked presentation conditions. *Cognition and Emotion*, *29*, 196–219.

Rosenberg, M. J., Hovland, C. I., McGuire, R. P., W. J.and Abelson, & Brehm, J. W. (1960). *Attitude organization and change: An analysis of consistency among attitude components*. New Haven, CA: Yale University Press.

Roskos-Ewoldsen, D. R., & Fazio, R. H. (1992). On the orienting value of attitudes: attitude accessibility as a determinant of an object's attraction of visual attention. *Journal of Personality and Social Psychology*, *63*, 198–211.

Sachisthal, M., Jansen, B., Peetsma, T., Dalege, J., van der Maas, H. L. J., & Raaijmakers, M. E. J. (2018). Introducing a science interest network model to reveal country differences. *Journal of Educational Psychology*. (Advance online publication)

Sayans-Jiménez, P., van Harreveld, F., Dalege, J., & Rojas Tejada, A. J. (2018). Investigating stereotype structure with empirical network models. *European Journal of Social Psychology*. (Advance online publication)

Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., ... Vandermeer, J. (2012). Anticipating critical transitions. *Science*, *338*, 344–348.

Schimmack, U., & Diener, E. (1997). Affect intensity: Separating intensity and frequency in repeatedly measured affect. *Journal of Personality and Social Psychology*, *73*, 1313–1329.

Schlicht-Schmälzle, R., Chykina, V., & Schmälzle, R. (2018). An attitude network analysis of post-national citizenship identities. *PloS one*, *13*, e0208241.

Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology*, *31*, 43–53.

Schneider, I. K., Eerland, A., van Harreveld, F., Rotteveel, M., van der Pligt, J., van der Stoep, N., & Zwaan, R. A. (2013). One way and the other: The bidirectional relationship between ambivalence and body movement. *Psychological Science*, *24*, 319–325.

Schrödinger, E. (1944). *What is life?* Cambridge: Cambridge University Press.

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition*, *25*, 638–656.

Shultz, T. R., & Lepper, M. R. (1996). Cognitive dissonance reduction as constraint satisfaction. *Journal of Personality and Social Psychology*, *103*, 219–240.

Siegler, R. S., Strauss, S., & Levin, I. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child development*, 1–84.

Simon, D., & Holyoak, K. J. (2002). Structural dynamics of cognition: From consistency theories to constraint satisfaction. *Personality and Social Psychology Review*, *6*, 283–294.

Simon, D., Krawczyk, D. C., & Holyoak, K. J. (2004). Construction of preferences by constraint satisfaction. *Psychological Science*, *15*, 331–336.

Simon, D., Pham, L. B., Le, Q. A., & Holyoak, K. J. (2001). The emergence of coherence over the course of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1250–1260.

Simon, D., Snow, C. J., & Read, S. J. (2004). The redux of cognitive consistency theories: Evidence judgments by constraint satisfaction. *Journal of Personality and Social Psychology*, *86*, 814–837.

Simon, D., & Spiller, S. A. (2016). The elasticity of preferences. *Psychological Science*, *27*, 1588–1599.

Simon, D., Stenstrom, D. M., & Read, S. J. (2015). The coherence effect: Blending cold and hot cognitions. *Journal of Personality and Social Psychology*, *109*, 369–394.

Simonton, D. K. (1999). *Origins of genius: Darwinian perspectives on creativity*. New York, NY: Oxford University Press.

Simonton, D. K. (2004). Psychology's status as a scientific discipline: Its empirical placement within an implicit hierarchy of the sciences. *Review of General Psychology, 8*, 59–67.

Sinclair, S., Lowery, B. S., Hardin, C. D., & Colangelo, A. (2005). Social tuning of automatic racial attitudes: the role of affiliative motivation. *Journal of Personality and Social Psychology, 89*, 583–592.

Sitnov, M. I., Sharma, A. S., Papadopoulos, K., & Vassiliadis, D. (2001). Modeling substorm dynamics of the magnetosphere: From self-organization and self-organized criticality to nonequilibrium phase transitions. *Physical Review E, 65*, 016116.

Smaldino, P. E. (2017). Models are stupid, and we need more of them. In R. R. Vallacher, J. Read, & A. Nowak (Eds.), *Computational social psychology* (pp. 311–331). New York, NY: Routledge.

Smith, M. B., Bruner, J. S., & White, R. W. (1956). *Opinions and personality*. Oxford, England: John Wiley & Sons.

Song, C., Qu, Z., Blumm, N., & Barabsi, A. L. (2010). Limits of predictability in human mobility. *Science, 327*, 1018–1021.

Sparks, P., Conner, M., James, R., Shepherd, R., & Povey, R. (2001). Ambivalence about health-related behaviours: An exploration in the domain of food choice. *British Journal of Health Psychology, 6*, 53–68.

Stapp, H. P. (1993). *Mind, matter, and quantum mechanics*. Berlin: Springer.

Stephan, K. E., Hilgetag, C. C., Burns, G. A., O'Neill, M. A., Young, M. P., & Kotter, R. (2000). Computational analysis of functional connectivity between areas of primate cerebral cortex. *Philosophical Transactions of the Royal Society, Series B: Biological Sciences, 355*, 111–126.

Strack, F., & Martin, L. (1987). Thinking, judging, and communicating: A process account of context effects in attitude surveys. In H. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 123–148). New York: Springer-Verlag.

Taleb, N. (2017). *How to forecast an election.* Retrieved from https://arxiv.org/abs/1703.06351

Tesser, A. (1978). Self-generated attitude change. *Advances in Experimental Social Psychology, 11*, 289–338.

Tesser, A. (1993). The importance of heritability in psychological research: The case of attitudes. *Psychological Review, 100*, 129–142.

Tesser, A., & Conlee, M. C. (1975). Some effects of time and thought on attitude polarization. *Journal of Personality and Social Psychology, 31*, 262–270.

Tesser, A., & Leone, C. (1977). Cognitive schemas and thought as determinants of attitude change. *Journal of Experimental Social Psychology, 13*, 340–356.

Tesser, A., Martin, L., & Mendolia, M. (1995). The impact of thought on attitude extremity and attitude-behavior consistency. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 73–92). Hillsdale, NJ: Lawrence Erlbaum.

Tesser, A., & Shaffer, D. R. (1990). Attitudes and attitude change. *Annual Review of Psychology, 41*, 479–523.

Thom, R. (1972). *Stabilitè structurelle et morphogenesè [Structural stability and morphogenesis]*. New York: Benjamin.

Thompson, M. M., Zanna, M. P., & Griffin, D. W. (1995). Let's not be indifferent about (attitudinal) ambivalence. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 361–386). Hillsdale, NJ: Lawrence Erlbaum.

Thurstone, L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273–286.

Tibshirani, R. (1996). Regression selection and shrinkage via the lasso. *Journal of the Royal Statistical Society B, 58*, 267–288.

Tononi, G. (2004). An information integration theory of consciousness. *BMC neuroscience, 5*, 42.

Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science, 282*, 1849–1851.

Torfs, P., & Bauer, C. (2014). *A (very short) introduction to R.* Retrieved from `https://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf`

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, England: Cambridge University Press.

Vallacher, R., R, & Nowak, A. (1997). The emergence of dynamical social psychology. *Psychological Inquiry, 8*, 73-99.

van Borkulo, C. D. (2016). *NetworkComparisonTest: Statistical comparison of two networks based on three invariance measures (R package version 2.0.1) [computer software].* Retrieved from `https://cran.r-project.org/web/packages/NetworkComparisonTest/index.html`

van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, B. W., Bosschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scientific Reports, 4*, 5918.

van Borkulo, C. D., Borsboom, D., Nivard, M., & Cramer, A. O. J. (2011). *Netlogo symptom spread model [computer software].* Retrieved from `http://ccl.northwestern.edu/netlogo/models/SymptomSpread` van Borkulo, C. D., Boschloo, L., Borsboom, D., Penninx, B. W. J. H., Waldorp, L. J., & Schoevers, R. A. (2015). Association of symptom network structure with the course of depression. *JAMA Psychiatry*, *72*, 1219–1226.

van Borkulo, C. D., & Epskamp, S. (2016). *Isingfit: Fitting ising models using the Elasso method (R package version 0.3.1) [computer software].* Retrieved from `https://cran.r-project.org/web/packages/IsingFit/index.html`

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Multivariate imputation by chained equations. *Journal of Statistical Software*, *45*, 1–67.

van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., ... Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 87–92.

van den Berg, H., Manstead, A. S. R., van der Pligt, J., & Wigboldus, D. H. J. (2005). The role of affect in attitudes toward organ donation and donor-relevant decisions. *Psychology and Health*, *20*, 789–802.

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842–861.

van der Maas, H. L. J., Kolstein, R., & van der Pligt, J. (2003). Sudden transitions in attitudes. *Sociological Methods & Research*, *2*, 125–152.

van der Pligt, J., de Vries, N. K., Manstead, A. S. R., & van Harreveld, F. (2000). The importance of being selective: Weighing the role of attribute importance. *Advances in Experimental Social Psychology*, *32*, 135–200.

Van Dessel, P., De Houwer, J., Hughes, S., & Hussey, I. (2018). An analysis of the scientific status and limitations of the attitudinal entropy framework and an initial test of some of its empirical predictions. *Psychological Inquiry*, *29*, 213-217.

van Geert, P., & van Dijk, M. (2015). The intrinsic dynamics of development. In R. A. Scott & S. M. Kosslyn (Eds.), . Wiley Online Library.

van Harreveld, F., Nohlen, H. U., & Schneider, I. K. (2015). The ABC of ambivalence: Affective, behavioral, and cognitive consequences of attitudinal conflict. *Advances in Experimental Social Psychology*, *52*, 285–324.

van Harreveld, F., Rutjens, B., Rotteveel, M., Nordgren, L., & van der Pligt,

J. (2009). Ambivalence and decisional conflict as a cause of psychological discomfort: Feeling tense before jumping off the fence. *Journal of Experimental Social Psychology*, *45*, 167–173.

van Harreveld, F., van der Pligt, J., & de Liver, Y. N. (2009). The agony of ambivalence and ways to resolve it: Introducing the MAID model. *Personality and Social Psychology Review*, *13*, 45–61.

van Harreveld, F., van der Pligt, J., de Vries, N. K., & Andreas, S. (2000). The structure of attitudes: attribute importance, accessibility and judgment. *British Journal of Social Psychology*, *39*, 363–380.

van Harreveld, F., van der Pligt, J., de Vries, N. K., Wenneker, C., & Verhue, D. (2004). Ambivalence and information integration in attititudinal judgment. *British Journal of Social Psychology*, *43*, 431–447.

van Overwalle, F., & Siebler, F. (2005). A connectionist model of attitude formation and change. *Personality and Social Psychology Review*, *9*, 231–274.

van Veen, V., Krug, M. K., Schooler, J. W., & Carter, C. S. (2009). Neural activity predicts attitude change in cognitive dissonance. *Nature Neuroscience*, *12*, 1469–1474.

Visser, P. S., Bizer, G., & Krosnick, J. A. (2006). Exploring the latent structure of strength-related attitude attributes. *Advances in Experimental Social Psychology*, *38*, 1–67.

Visser, P. S., & Krosnick, J. A. (1998). Development of attitude strength over the life cycle: surge and decline. *Journal of Personality and Social Psychology*, *75*, 1389–1410.

Visser, P. S., Krosnick, J. A., & Simmons, J. P. (2003). Distinguishing the cognitive and behavioral consequences of attitude importance and certainty: A new approach to testing the common-factor hypothesis. *Journal of Experimental Social Psychology*, *39*, 118–141.

Visser, P. S., & Mirabile, R. R. (2004). Attitudes in the social context: the impact of social network composition on individual-level attitude strength. *Journal of Personality and Social Psychology*, *87*, 779–795.

Wainwright, M. J., & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends®in Machine Learning*, *1*, 1–305.

Wallis, W. D. (2007). *A beginner's guide to graph theory*. New York, NY: Birkhäuser.

Wang, Z., & Busemeyer, J. R. (2013). A quantum question order model supported by empirical tests of an a priori and precise prediction. *Topics in Cognitive Science*, *5*, 689–710.

Wang, Z., Solloway, T., Shiffrin, R., & Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of sciences of the United States of America*, *111*, 9431–9436.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Watts, D. J. (1999). Networks, dynamics, and the small-world phenomen. *American Journal of Sociology*, *105*, 493–527.

Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 5766–5771.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440–442.

West, D. B. (1996). *Introduction to graph theory*. Upper Saddle River, NJ: Prentice Hall.

Westaby, J. D., Pfaff, D. L., & Redding, N. (2014). Psychology and social networks: A dynamic network theory perspective. *American Psycholo- gist*, *69*, 269–284.

Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues*, *25*, 41–78.

Wilensky, U. (1999). *Netlogo*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wilson, T. D., & Hodges, S. D. (1992). Attitudes as temporary constructions. In M. L. L. & A. Tesser (Eds.), *The construction of social judgments* (pp. 37–65). Hillsdale, NJ: Lawrence Erlbaum.

Wlezien, C. (2003). Presidential election polls in 2000: A study in dynamics. *Presidential Studies Quarterly*, *33*, 172–186.

Wood, C., Conner, M., Miles, E., Sandberg, T., Taylor, N., Godin, G., & Sheeran, P. (2016). The impact of asking intention or self-prediction questions on subsequent behavior: a meta-analysis. *Personality and Social Psychology Review*, *20*, 245–268.

Wood, W., Rhodes, N., & Biek, M. (1995). Working knowledge and attitude strength: An information-processing analysis. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 283–314). Hillsdale, NJ: Lawrence Erlbaum.

Wu, F. Y. (1982). The potts model. *Reviews of Modern Physics*, *54*, 235–268.

Young, G., & Smith, R. (2005). *Essentials of statistical inference*. Cambridge, UK: Cambridge University Press.

Zald, D. H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain Research Reviews*, *41*, 88–123.

Zanna, M. P., & Cooper, J. (1974). Dissonance and the pill: An attribution approach to studying the arousal properties of dissonance. *Journal of Personality and Social Psychology*, *29*, 703–709.

Zanna, M. P., & Rempel, J. K. (1988). Attitudes: A new look at an old concept. In D. Bar-Tal & A. W. Kruglanski (Eds.), *The psychology of knowledge* (pp. 315–334). Cambridge, UK: Cambridge University Press.

Zeeman, E. C. (1976). Catastrophe theory. *Scientific American*, *234*, 65–83.

Appendix A

Description of the ANES 1984 Data Set

A.1 Evaluative Reactions

Sixteen evaluative reactions toward the presidential candidates tapping beliefs and seven evaluative reactions tapping feelings were assessed. For items tapping beliefs, participants were asked: "In your opinion, does the phrase 'he...' describe the candidate?". The beliefs that completed the items were "is moral", "is knowledgeable", "is inspiring", "would provide strong leadership", "is hard-working", "is decent", "is compassionate", "commands respect", "is intelligent", "is kind", "sets a good example", "really cares about people like you", "understands people like you", "is fair", "is in touch with ordinary people", and "is religious". We excluded the item focusing on whether the participants believed that the candidates were religious because this belief is ambiguous in relation to valence. Items were assessed on a 4-point scale, with answer options 4 = "Extremely well", 3 = "Quite well", 2 = "Not too well", 1 = "Not well at all". As the eLasso-procedure only allows dichotomous data (van Borkulo et al., 2014), we scored option 3 and 4 as 1 and option 1 and 2 as 0.

For items tapping feelings, participants were asked: "Has the candidate – because of the kind of person he is or because of something he has done, ever made you feel: ...?". The feelings that completed the items were "angry", "hopeful", "afraid of him", "proud", "disgusted", "sympathetic toward him", and "uneasy". These items were assessed dichotomously, with the answer options 1 = "Yes" and 0 = "No".

A.2 Missing Data

Some participants had missing data on at least one of the evaluative reactions because they responded "Don't know". We deleted missing data casewise, which let to the final sample of 1877 participants for the analysis on the attitude toward Ronald Reagan and 1628 participants for the analysis on the attitude toward Walter Mondale. To check the robustness of our results, we also imputed missing values with randomly assigning 0 or 1, when a participant had missing values. The results of this analysis mirrored the results reported here.

Appendix B

Supplementary Information on Chapter 3

B.1 Estimating Networks from Continuous or Categorical Data

For estimating networks from continuous or categorical data, we can use the function `EBICglasso` in combination with the function `cor_auto`. Both functions are provided in the R package *qgraph* (Epskamp, Costantini, et al., 2018). `EBICglasso` is used in a similar way as the `IsingFit` function, but requires as input a correlation matrix instead of a data frame. For this we can use the function `cor_auto`, which detects whether variables are continuous or categorical and then calculates the appropriate correlation matrix (e.g., polychorious correlations if the variables are categorical). If our data were continuous or categorical, we could estimate attitude networks in the following way:

```
ObamaWeiAdj <- EBICglasso(cor_auto(Obama), nrow(Obama))
```

This command estimates a network from the data and stores its weight adjacency matrix to the object `ObamaWeiAdj`. The first argument provides the input correlation matrix and the second argument specifies the number of cases on which the correlation matrix was estimated. We can then use the weight adjacency matrix as input for a *qgraph* object or an *igraph* object and run the same analyses on this network as we described in the tutorial.

```
ObamaqGraph2 <- qgraph(ObamaWeiAdj)
ObamaiGraph2 <- graph_from_adjacency_matrix(abs(ObamaWeiAdj),
    ↪  'undirected', weighted = TRUE, add.colnames = FALSE)
```

B.2 Inclusion of Covariates

It is also possible to add covariates when one wants to estimate an attitude network. We would advice researches, who want to do this, to first estimate a network including the relevant attitude items and the covariates. Second, one can extract the network based only on the attitude items and run network analyses on this network and plot the network without the covariates. Adding the covariates gender and age to the Obama network can be done in this way:

```
ObamaCov <- data.frame(ObamaCog, ObamaAff, gender, age)
ObamaCov <- na.omit(ObamaCov)
ObamaFitCov <- mgm(ObamaCov, c(rep('c', 11),'g'), c(rep(2,
    ↪ 11),1), binarySign = TRUE)
```

Note that we use the function `mgm`, available in the R package *mgm* (Haslbeck, 2017), because we need to estimate a network from data involving different kind of variables, as the evaluative reactions and gender are binary variables and age is a continuous variable. The first argument in the function specifies the data frame on which the network is estimated, the second argument specifies the type of the variables, with `'c'` indicating categorical variables and `'g'` indicating Gaussian or continuous variables, and the `binarySign` argument specifies that we also want to estimate weather binary variables are positively or negatively connected (which is not done by default in the `mgm` function). To get a weight adjacency matrix similar to the weight adjacency matrix of the `IsingFit` function, we need to combine the following two outputs of the `mgmfit` function:

```
ObamaFitCov$pairwise$signs[is.na(ObamaFitCov$pairwise$signs)]
    ↪  <- 0
ObamaNetCovIn <- ObamaFitCov$pairwise$wadj*ObamaFitCov$
    ↪ pairwise$signs
```

The object `ObamaNetCovIn` now contains the weight adjacency matrix of the evaluative reactions and the covariates. We can select the weight adjacency matrix of only the evaluative reactions in the following way:

```
ObamaNetCovOut <- ObamaNetCovIn[1:10,1:10]
```

The object `ObamaNetCovOut` now contains the weight adjacency matrix of the evaluative reactions that is estimated when gender and age is controlled for.

B.3 Edge Stability

As can be seen in Figure 3.2, edges differ quite strongly in their magnitude. If one assumes that the edges represent causal connections, one can use the magnitude of the edges to infer to what extent change in one evaluative reaction would cause change in another evaluative reaction. However, edges are subject to sampling error, so that, to be able to interpret these differences, we need to assess the stability of the network structure (Epskamp, Borsboom, & Fried, 2018). To assess the stability of the edge weights, we can use the function `bootnet`, available in the R package *bootnet* (Epskamp, 2017). This function creates samples based on bootstrapping and returns estimates of how strongly network parameters differ between samples. We can run the function `bootnet` on the Obama network in this way:

```
ObamaSta <- bootnet(Obama, 1000, 'IsingFit')
```

The first argument of the `bootnet` function specifies the data on which the function is run, the second argument specifies the number of bootstrapped samples, and the third argument specifies the estimation method for creating the networks. We can plot the results of the bootstrapping analysis for the stability of the edge weights in this way:

```
plot(ObamaSta, plot = 'interval', order = 'sample')
```

This function creates the plot shown in Figure B.1. The first argument in the plot function specifies that we want to plot the object `ObamaSta`, the `plot` argument specifies that we want to plot the bootstrapped confidence intervals around the edge weights and the `order` argument specifies that we want the edges to be ordered according to their magnitude.

As can be seen in Figure B.1, the confidence intervals are relatively small, which indicates that large differences in edge weights are meaningful. The strongest positive edge is the connection between judging Barack Obama as knowledgeable and as intelligent (Kno–Int). The confidence interval of this edge only overlaps with the confidence interval of the second strongest positive edge, which is the connection between feeling hopeful and feeling proud toward Obama (Hop–Prd). It is thus certain that the edge Know-Int differs significantly from all the other edges except from the edge Hope–Prou. The edges Know–Int and Hope–Prou might also differ significantly, as their confidence intervals overlap only slightly. Edges that have largely overlapping confidence intervals should not be interpreted as differing in their magnitude.

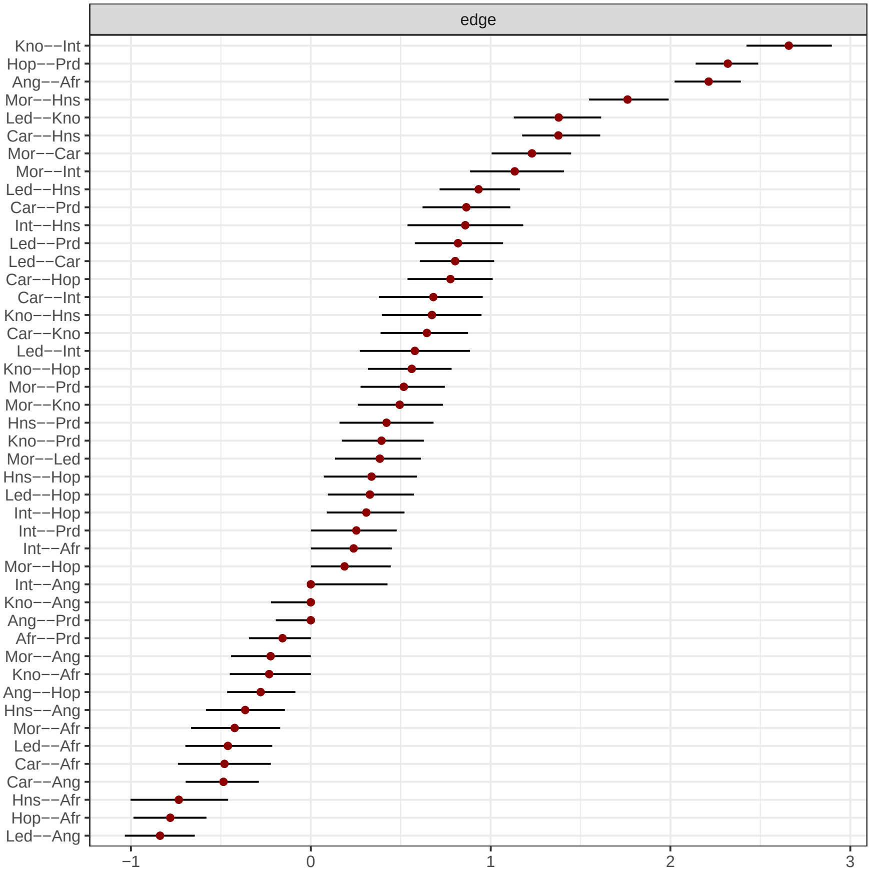


Figure B.1: Plot of the stability of the edge estimates of the Obama network. Points indicate the mean edge weights of the bootstrapped samples and lines indicate the width of the 95% confidence interval. See Table 3.1 for the abbreviations of the nodes.

B.4 Illustration of Shortest Path Lengths

To make the concept of shortest path lengths less abstract we use two examples as illustrations. As a first example, take the shortest path length (l) between the nodes Kno and Int, which is equal to the inverse of their direct strong connection, because no indirect connection between these two nodes represents a shorter path ($l_{Kno-Int}$ = 2.66-1 = 0.38). Conversely, the shortest path between the nodes Led and Int does not reflect the direct connection between these two nodes, as the direct connection between the two nodes is weak. The shortest path between these two nodes runs through the node Kno. The shortest path length between Led and Int then is the sum of the inverses of the edge weights between the nodes Kno and Int and between the nodes Led and Kno ($l_{Led-Int}$ = 1.38-1 + 2.66-1 = 1.10), which is smaller than the path based on the direct connection between Led and Int.

B.5 Centrality Stability

Centrality indices are relatively complicated functions of a potentially large number of statistical estimates, each of which is subject to sampling error. As a result, these indices can vary quite substantially across samples and their values should be interpreted with care. Fortunately, we can again use the bootnet function to investigate the stability of the different centrality indices. In contrast to assessing the stability of the edges, we cannot use bootstrapping to assess the stability of the centrality indices (Epskamp, 2017). An alternative to bootstrapping is to investigate how strongly the centrality estimates are affected when subsets of the sample are used to estimate the network. We can specify the bootnet function to do this in the following way:

```
ObamaCenSta <- bootnet(Obama, 1000, 'IsingFit', 'person')
```

Setting the fourth argument of the bootnet function to 'person' results in the calculation of the centrality estimates for different subsamples of the data. We can plot the results of this analysis in the following way:

```
plot(ObamaCenSta, subsetRange = c(100,50))
```

This command creates the plot shown in Figure B.2. The argument subsetRange specifies the samples we want the plot to show. In this case the plot shows the correlation between the centrality estimates of the whole sample and samples based on 50 to 100 % of the data. As can be seen in Figure B.2, both the strength and closeness estimates are highly stable. Even for the samples including only 50 % of the individuals of the complete sample, the correlation with the centrality indices of the complete sample stays close to 1. It is thus safe to interpret the estimated strength and closeness estimates of the Obama network. For the betweenness estimates the story is a bit different as the correlation drops continuously when less individuals are included. The correlation, however, stays positive and also of relatively high magnitude. To further investigate the stability of the betweenness estimates, we can plot how much each betweeness estimate fluctuates in the different samples. We can do this by using this command:

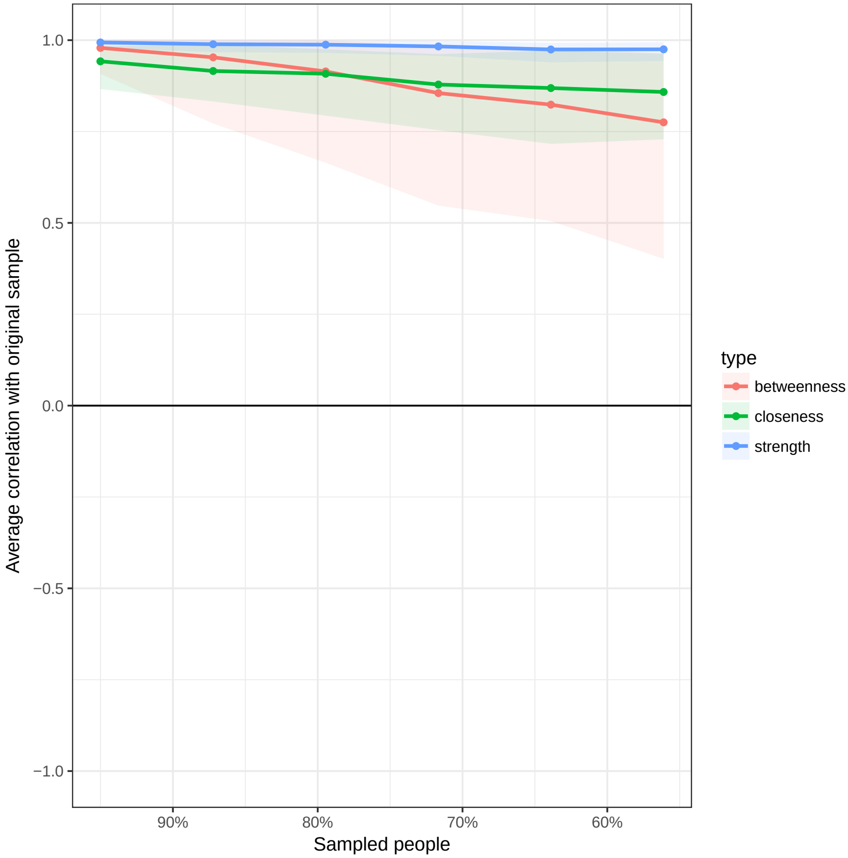


Figure B.2: Stability plot of the centrality estimates of the Obama network. The lines represent the mean correlations between the given sampled percentage and the complete sample and the shades represent the area between the 2.5% and 97.5% percentile of the sampled estimates.

```
plot(ObamaCenSta, 'betweenness', perNode = TRUE, subsetRange
    ↪ = c(100,50))
```

This command creates the plot in Figure B.3. The second argument of the plot function specifies that we only want to plot the betweeness estimates and the perNode argument specifies that we want to plot the betweenness estimates of every node. As can be seen in Figure B.3, the ordering of the betweeness estimates is sufficiently stable: Only nodes that have the same betweeness estimates in the complete sample change their position in the different samples. It is thus also safe to interpret the betweenness estimates of the Obama network.

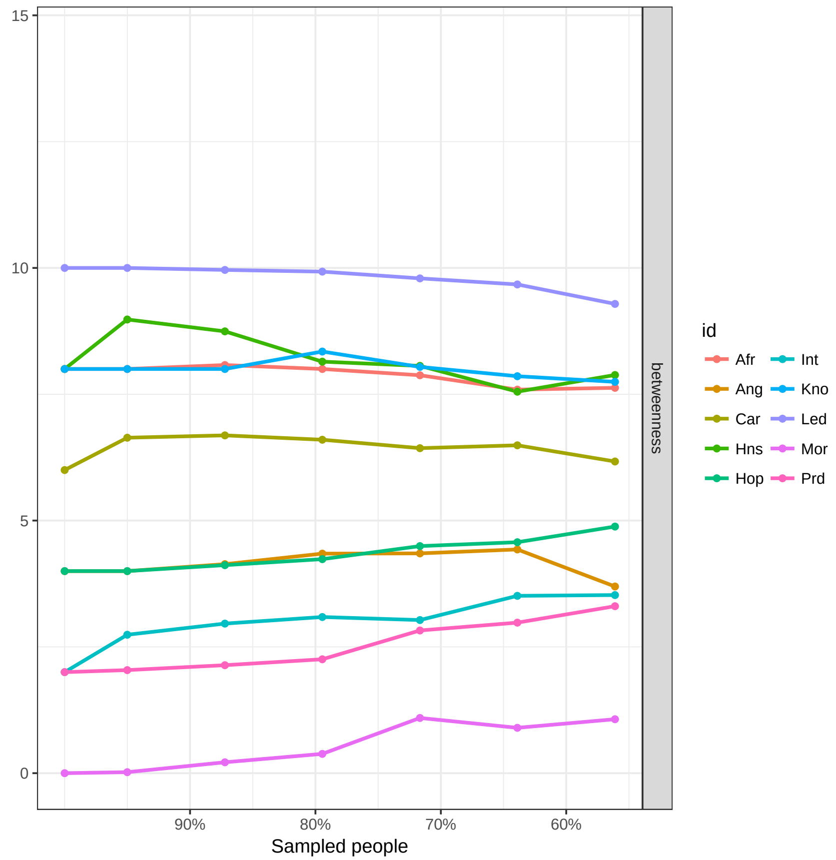


Figure B.3: Stability plot of the betweenness estimates of the Obama network. See Table 3.1 for the abbreviations of the nodes.

Often results of the stability analyses might not be as positive as in our current example. We would advise researchers to first test whether the correlations between the centrality estimates of the complete sample and the subsamples stay close to 1 (as is the case here for the closeness and strength indices). If this turns out to be the case it is safe to interpret the ordering of the centrality indices. In the case that correlations clearly drop below 1 (as is the case here for the betweenness indices), we advice researchers to investigate whether the ordering of the centrality indices remains the same for the different subsamples. Differences between indices that change their position in the different subsamples should not be interpreted, but one can still interpret the differences between the indices that do not change their position (e.g., one node stands out as highest in betweenness while the other nodes change their positions).

Appendix C

Supplementary R Code to Chapter 3

```
#########################
#Install packages (uncomment if you have not installed them already)

#install.packages("foreign")
#install.packages("IsingFit")
#install.packages("qgraph")
#install.packages("igraph")
#install.packages("bootnet")
#install.packages("IsingSampler")
#install.packages("compute.es")
#install.packages("NetworkComparisonTest")
#install.packages("mgm")
#install.packages("Matrix")

#########################
#Load packages

library(foreign)
library(IsingFit)
library(qgraph)
library(igraph)
library(bootnet)
library(IsingSampler)
library(compute.es)
library(NetworkComparisonTest)
library(mgm)
library(Matrix)

#########################
#Create example network

set.seed(1)#this is used so that each time the same random numbers are
    ↪ used

exNet <- sample_pa(10, .5, m = 2, directed = FALSE)#this creates a network
    ↪  with ten nodes and 17 edges based on the growth algorithm
    ↪ preferential attachment
```

```
pdf('Figure1.pdf')#the pdf function is used to save the plot to a pdf file
qgraph (get.adjacency(exNet))#this plots the network
dev.off()

########################
#Load data

download.file('http://www.electionstudies.org/studypages/data/anes_
    ↪ timeseries_2012/anes_timeseries_2012_dta.zip', 'ANES2012.zip') #
    ↪ downloads the zipped data file to your working directory
unzip('ANES2012.zip')
ANES2012 <- read.dta('anes_timeseries_2012_Stata12.dta')#loads the data to
    ↪  the object ANES2012

########################
#Recode variables
#Items regarding Obama

ObamaCog <- data.frame(Mor = as.numeric(ANES2012$ctrait_dpcmoral),#this
    ↪ creates a data frame containing the items tapping beliefs
                    Led = as.numeric(ANES2012 $ ctrait_dpclead),
                    Car = as.numeric(ANES2012$ctrait_dpccare),
                    Kno = as.numeric(ANES2012$ctrait_dpcknow),
                    Int = as.numeric(ANES2012$ctrait_dpcint),
                    Hns = as.numeric(ANES2012$ctrait_dpchonst))
ObamaCog[ObamaCog < 3] <- NA#values below 3 represent missing values
ObamaCog <- binarize(ObamaCog, 5, removeNArows = FALSE)#this binarizes the
    ↪  data (this is done because the model we use for simulating
    ↪ networks assumes binary data); (not) endorsing the beliefs is
    ↪ encoded as 1 (0)

ObamaAff <- data.frame(Ang = as.numeric(ANES2012$candaff_angdpc),#this
    ↪ creates a data frame containing the items tapping feelings
                    Hop = as.numeric(ANES2012$candaff_hpdpc),
                    Afr = as.numeric(ANES2012$candaff_afrdpc),
                    Prd = as.numeric(ANES2012$candaff_prddpc))
ObamaAff[ObamaAff < 3] <- NA#values below 3 represent missing values
ObamaAff <- binarize(ObamaAff, 4, removeNArows = FALSE)#(not) endorsing
    ↪ the feelings is encoded as 1 (0)

Obama <- data.frame(ObamaCog,ObamaAff)#this creates a data frame
    ↪ containing all items tapping evaluative reactions
Obama <- na.omit(Obama)#this deletes missing values casewise

#Items regarding Romney

RomneyCog <- data.frame(Mor = as.numeric(ANES2012$ctrait_rpcmoral),
                    Led = as.numeric(ANES2012 $ ctrait_rpclead),
                    Car = as.numeric(ANES2012$ctrait_rpccare),
```

```
                    Kno = as.numeric(ANES2012$ctrait_rpcknow),
                    Int = as.numeric(ANES2012$ctrait_rpcint),
                    Hns = as.numeric(ANES2012$ctrait_rpchonst))
RomneyCog[RomneyCog < 3] <- NA
RomneyCog <- binarize(RomneyCog, 5, removeNArows = FALSE)
RomneyAff <- data.frame(Ang = as.numeric(ANES2012$candaff_angrpc),
                    Hop = as.numeric(ANES2012$candaff_hprpc),
                    Afr = as.numeric(ANES2012$candaff_afrrpc),
                    Prd = as.numeric(ANES2012$candaff_prdrpc))
RomneyAff[RomneyAff < 3] <- NA
RomneyAff <- binarize(RomneyAff, 4, removeNArows = FALSE)

Romney <- data.frame(RomneyCog,RomneyAff)

#########################
#Network estimation

ObamaFit <- IsingFit(Obama)
ObamaGraph <- qgraph(ObamaFit $ weiadj, layout = 'spring', cut = .8)
ObamaiGraph <- graph_from_adjacency_matrix(abs(ObamaFit $ weiadj), '
    ↪ undirected', weighted = TRUE, add.colnames = FALSE)

#########################
#Community detection and plotting

ObamaCom <- cluster_walktrap(ObamaiGraph)
communities(ObamaCom)

pdf('Figure2.pdf')
qgraph(ObamaFit $ weiadj, layout = 'spring', cut = .8, groups =
    ↪ communities(ObamaCom), legend = FALSE)
dev.off()

#plotting the network in a colorblind-friendly way
qgraph(ObamaFit $ weiadj, layout = 'spring', cut = .8, groups =
    ↪ communities(ObamaCom), legend = FALSE,
      theme = 'colorblind')#this argument can be used to plot the network
          ↪ in way that is more readable for colorbling people

#plotting the network in greyscale
qgraph(ObamaFit $ weiadj, layout = 'spring', cut = .8, groups =
    ↪ communities(ObamaCom), legend = FALSE,
      theme = 'gray')#this argument can be used to plot the network in
          ↪ greyscale



#########################
#Centrality
```

```
ObamaCen <- centralityTable(ObamaGraph, standardized = FALSE)

pdf('Figure3.pdf')
centralityPlot(ObamaGraph, scale = 'raw')
dev.off()

#########################
#Connectivity

ObamaSPL <- centrality(ObamaGraph)$ShortestPathLengths
ObamaSPL <- ObamaSPL[upper.tri(ObamaSPL)]
ObamaASPL <- mean(ObamaSPL)

#########################
#Comparison between Obama and Romney network

ObaRom <- data.frame(ObamaCog,ObamaAff,Romney)#this creates a data frame
    ↪ containing all items regarding Obama and Romney
ObaRom <- na.omit(ObaRom)#this is done so that only individuals, who have
    ↪ no missing values on items regarding both Obama and Romney, are
    ↪ included

ObamaComp <- ObaRom[,1:10]#this selects the items regarding Obama
RomneyComp <- ObaRom[,11:20]#this selects the items regarding Romney
names(RomneyComp) <- names(ObamaComp)#this makes sure that both data
    ↪ frames have the same names (which is neccessary for the NCT)

set.seed(1)

NCTObaRom <- NCT(ObamaComp, RomneyComp, it = 1000, binary.data = TRUE,
    ↪ paired = TRUE, test.edges = TRUE, edges = 'all')
NCTObaRom$glstrinv.real
NCTObaRom$glstrinv.pval
NCTObaRom$nwinv.real
NCTObaRom$nwinv.pval
NCTObaRom$einv.pvals

#input for Figure 4
ObamaCompFit <- IsingFit(ObamaComp)#this creates the Obama network that
    ↪ was tested in the NCT
RomneyCompFit <- IsingFit(RomneyComp)#this creates the Romney network that
    ↪  was tested in the NCT

inputNCTgraph <- ObamaCompFit$weiadj - RomneyCompFit$weiadj#this creates
    ↪ the input for Figure 4, which is based on the differences in edge
    ↪ weights between the Obama and Romney networks
inputNCTgraph[upper.tri(inputNCTgraph)][which(NCTObaRom$einv.pvals$`p-
    ↪ value` >= .05)] <- 0#this sets the not-significant edge weight
    ↪ differences in the upper triangle of the input matrix to 0
```

```
inputNCTgraph <- forceSymmetric(inputNCTgraph)#this sets the lower
    ↪ triangle of the matrix to be symmetric with the upper triangle

pdf('Figure4.pdf')
qgraph(inputNCTgraph, layout = ObamaGraph$layout, edge.labels = TRUE,
    ↪ esize = 1)#the layout argument specifies that we want the same
    ↪ layout as in the Obama graph, the edge.labels argument specifies
    ↪ that we want to plot the values of the edge weights, esize reduces
    ↪ the width of the edges
dev.off()

#########################
#Simulation input

SimInput <- LinTransform(ObamaFit$weiadj, ObamaFit$thresholds)#The
    ↪ LinTransform function rescales the edge weights and thresholds, so
    ↪ that a threshold of 0 indicates that a node has no disposition to
    ↪ be in a given state and a positive (negative) threshold indicates
    ↪ that a node has the disposition to be "on" ("off")
SimInput$graph[c(7,9),-c(7,9)] <- SimInput$graph[c(7,9),-c(7,9)]*-1#this
    ↪ command and the next command rescale the nodes Ang and Afr
SimInput$graph[-c(7,9),c(7,9)] <- SimInput$graph[-c(7,9),c(7,9)]*-1

#########################
#Connectivity simulation

set.seed(1)

sampleLowTemp <- IsingSampler(1000, SimInput $ graph, rep(0,10), 1.2,
    ↪ responses = c(-1L,1L))
sampleMidTemp <- IsingSampler(1000, SimInput $ graph, rep(0,10), .8,
    ↪ responses = c(-1L,1L))
sampleHighTemp <- IsingSampler(1000, SimInput $ graph, rep(0,10), .4,
    ↪ responses = c(-1L,1L))

#the simulated data contains -1;+1 responses. To estimate networks from
    ↪ the simulated data using IsingFit, 0s have to be assigned -1
    ↪ responses

sampleLowTempResc <- sampleLowTemp
sampleLowTempResc[sampleLowTempResc == -1] <- 0

sampleMidTempResc <- sampleMidTemp
sampleMidTempResc[sampleMidTempResc == -1] <- 0

sampleHighTempResc <- sampleHighTemp
sampleHighTempResc[sampleHighTempResc == -1] <- 0

#these commands fit networks on the simulated data
sampleLowTempFit <- IsingFit(sampleLowTempResc)
```

```
sampleMidTempFit <- IsingFit(sampleMidTempResc)
sampleHighTempFit <- IsingFit(sampleHighTempResc)

pdf('Figure5.pdf', 7, 10.5)#this creates Figure 5
layout(matrix( c(1, 2,
                 3, 4,
                 5, 6), 3, 2, byrow = TRUE))

qgraph(sampleHighTempFit $ weiadj, layout = ObamaGraph $ layout, maximum =
    ↪  max(abs(sampleLowTempFit $ weiadj)),
      cut = quantile(abs(sampleMidTempFit $ weiadj), .75), label.font = 2)
hist(apply(sampleHighTemp, 1, sum), breaks = seq(-11,11,2), include.lowest
    ↪  = FALSE, axes = FALSE,
    xlab = 'Sum␣Score', main = '')
axis(1, seq(-10,10,2), seq(-10,10,2), cex.axis = .9)
axis(2)
mtext('High␣Temperature', line = 2, at = -17, font = 2)

qgraph(sampleMidTempFit $ weiadj, layout = ObamaGraph $ layout, maximum =
    ↪ max(abs(sampleLowTempFit $ weiadj)),
      cut = quantile(abs(sampleMidTempFit $ weiadj), .75), label.font = 2)
hist(apply(sampleMidTemp, 1, sum), breaks = seq(-11,11,2), include.lowest
    ↪ = FALSE, axes = FALSE,
     xlab = 'Sum␣Score', main = '')
axis(1, seq(-10,10,2), seq(-10,10,2), cex.axis = .9)
axis(2)
mtext('Mid␣Temperature', line = 2, at = -17, font = 2)

qgraph(sampleLowTempFit $ weiadj, layout = ObamaGraph $ layout, maximum =
    ↪ max(abs(sampleLowTempFit $ weiadj)), #the layout argument specifies
    ↪  that all networks are plotted in the layout of the Obama network;
    ↪ the maximum argument makes the different networks comparable as the
    ↪  maximum value is set to the highest edge in the low temperature
    ↪ network
      cut = quantile(abs(sampleMidTempFit $ weiadj), .75), label.font = 2)
          ↪ #the cut argument is used to plot only edges higher than the
          ↪ 75% qunatile of all networks with higher width
hist(apply(sampleLowTemp, 1, sum), breaks = seq(-11,11,2), include.lowest
    ↪ = FALSE, axes = FALSE,
    xlab = 'Sum␣Score', main = '')
axis(1, seq(-10,10,2), seq(-10,10,2), cex.axis = .9)
axis(2)
mtext('Low␣Temperature', line = 2, at = -17, font = 2)

dev.off()

#########################
#Centrality simulation

set.seed (1)
```

```
SampleNeg <- IsingSampler(1000, SimInput$graph,
                      rep(-.1,10),#this argument specifies the thresholds,
                          ↪  which are all set to -.1
                      responses = c(-1L,1L))
SampleHns <- IsingSampler(1000, SimInput$graph,
                      c(rep(-.1,5),#this sets the thesholds of the first
                          ↪ five nodes to -.1
                        1,#this sets the threshold of the sixth node,
                            ↪ which is the node Hns, to 1
                        rep(-.1,4)),#this sets the thesholds of the last
                            ↪ four nodes to -.1
                      responses = c(-1L,1L))
SampleAng <- IsingSampler (1000, SimInput$graph, c(rep(-.1,6),#this sets
    ↪ the thesholds of the first six nodes to -.1
                                           1,#this sets the threshold of
                                               ↪ the seventh node, which
                                               ↪  is the node Ang, to 1
                                           rep(-.1,3)),#this sets the
                                               ↪ thesholds of the last
                                               ↪ three nodes to -.1
                       responses = c(-1L,1L))

#calculate the sum scores of the different networks
sumSampleNeg <- apply(SampleNeg, 1, sum)
sumSampleHns <- apply(SampleHns, 1, sum)
sumSampleAng <- apply(SampleAng, 1, sum)

#calculate descriptive statistics of the different networks
mean(sumSampleNeg)
sd(sumSampleNeg)

mean(sumSampleHns)
sd(sumSampleHns)

mean(sumSampleAng)
sd(sumSampleAng)

#perform t-test on the sum scores of the different networks
t.test(sumSampleHns, sumSampleAng, var.equal = TRUE)

#calculate the effect size of the difference in the sum scores
mes(mean(sumSampleHns), mean(sumSampleAng), sd(sumSampleHns), sd(
    ↪ sumSampleAng), 1000, 1000)

#########################
#Supplementary information

#########################
#Estimating networks based on continious (or categorical) data
```

```
ObamaWeiAdj <- EBICglasso(cor_auto(Obama), nrow(Obama))
ObamaqGraph2 <- qgraph(ObamaWeiAdj)
ObamaiGraph2 <- graph_from_adjacency_matrix(abs(ObamaWeiAdj), 'undirected'
    ↪ , weighted = TRUE, add.colnames = FALSE)

#########################
#Covariates

gender <- as.numeric(ANES2012$gender_respondent_x)#this saves the gender
    ↪ of the participants to the numeric verctor gender
gender <- gender-1#this rescores the values into 0=male and 1=female
age <- ANES2012$dem_age_r_x#this saves the age of the participants to the
    ↪ numeric verctor age
age[age < 1] <- NA#values lower than 1 indicate missing values

ObamaCov <- data.frame(ObamaCog, ObamaAff, gender, age)
ObamaCov <- na.omit(ObamaCov)
ObamaFitCov <- mgm(ObamaCov, c(rep('c', 11),'g'), c(rep(2, 11),1),
    ↪ binarySign = TRUE)
ObamaFitCov$pairwise$signs[is.na(ObamaFitCov$pairwise$signs)] <- 0
ObamaNetCovIn <- ObamaFitCov$pairwise$wadj*ObamaFitCov$pairwise$signs
ObamaNetCovOut <- ObamaNetCovIn[1:10,1:10]

#########################
#Edge stability

set.seed(1)

ObamaSta <- bootnet(Obama, 1000, 'IsingFit')

pdf('FigureS1.pdf')
plot(ObamaSta, plot = 'interval', order = 'sample')
dev.off()

#########################
#Centrality Stability

set.seed(1)

ObamaCenSta <- bootnet(Obama, 1000, 'IsingFit', 'person')

pdf('FigureS2.pdf')
plot(ObamaCenSta, subsetRange = c(100,50))
dev.off()

pdf('FigureS3.pdf')
plot(ObamaCenSta, c('betweenness'), perNode = TRUE, subsetRange = c
    ↪ (100,50))
dev.off()
```

Appendix D

Supplementary Notes on Chapter 5

D.1 Supplementary Note 1: Analytical Solutions of the Hypotheses

In the Ising model the intuition is that higher connection strength will allow for better prediction. Here we show this intuition is correct.

D.1.1 Logistic Regression and the Ising Model

The Ising model is part of the exponential family of distributions (Brown, 1986; Young & Smith, 2005; Wainwright & Jordan, 2008). Let G be a graph consisting of nodes in $V = \{1, 2, \ldots, p\}$ and edges (s, t) in $E \subseteq V \times V$. To each node $s \in V$ a random variable X_s is associated with values in $\{0, 1\}$. The probability of each configuration x depends on a main effect (external field) and pairwise interactions. It is sometimes referred to as the auto logistic-function (Besag, 1974), or a pairwise Markov random field, to emphasise that the parameter and sufficient statistic space are limited to pairwise interactions (Wainwright & Jordan, 2008). Each $x_s \in \{0, 1\}$ has conditional on all remaining variables (nodes) $X_{\backslash s}$ probability of success $\pi_s := \mathbb{P}(X_s = 1 \mid x_{\backslash s})$. The distribution for configuration x of the Ising model is then

$$\mathbb{P}(x) = \frac{1}{Z(\theta)} \exp\left(\sum_{s \in V} m_s x_s + \sum_{(s,t) \in E} A_{st} x_s x_t \right) \tag{D.1}$$

which is clearly of the form of exponential family. In general, the normalisation $Z(\theta)$ is intractable, because the sum consists of 2^p possible configurations for $y \in \{0, 1\}^p$; for example, for $p = 30$ we obtain over 1 million configurations to evaluate in the sum in $Z(\theta)$ (see Wainwright & Jordan, 2008 for lattice [Bethe] approximations). The conditional distribution is again an Ising model (Besag, 1974; Kolaczyk, 2009)

$$\pi_s = \mathbb{P}(x_s = 1 \mid x_{\setminus s}) = \frac{\exp\left(m_s + \sum_{t:(s,t)\in E} A_{st}x_t\right)}{1 + \exp\left(m_s + \sum_{t:(s,t)\in E} A_{st}x_t\right)}. \tag{D.2}$$

It immediately follows that the log-odds (Besag, 1974) is

$$\mu_s(x_{\setminus s}) = \log\left(\frac{\pi_s}{1-\pi_s}\right) = m_s + \sum_{t:(s,t)\in E} A_{st}x_t. \tag{D.3}$$

Note that the log-odds $\theta \mapsto \mu_\theta$ is a linear function, and so if $x = (1, x_{\setminus s})$ then $\mu_\theta = x^\mathsf{T}\theta$. Recall that $\theta \mapsto \mu_\theta$ is the linear function $\mu_{\theta_s}(x_{\setminus s}) = m_s + \sum_{t\in V\setminus s} A_{st}x_t$ of the conditional Ising model obtained from the log-odds (D.3). Define $\mu_s := \mu_{\theta_s}(x_{\setminus s})$. We use the notation that the node of interest $x_{i,s}$ is denoted by y_i and we let the remaining variables and a 1 for the intercept be indicated by $x_i = (1, x_{i,\setminus s})$, basically leaving out the subscript s to index the node, and only use it whenever circumstances demand it. Let the loss function be the negative log of the conditional probability π in (D.2), known as a pseudo log-likelihood (Besag, 1974)

$$\psi(x,\mu) := -\log \mathbb{P}(y \mid x) = -x\mu + \log(1 + \exp(\mu)). \tag{D.4}$$

D.1.2 Monotonicity of Prediction Loss as a Function of Connectivity

In logistic regression there is a natural classifier that predicts whether y_i is 1 or 0. We simply check whether the probability of a 1 is greater than 1/2, that is, whether $\pi_i > 1/2$. Because $\mu_i > 0$ if and only if $\pi_i > 1/2$ we obtain the natural classifier

$$\mathcal{C}(y_i) = \mathbb{1}\{\mu_i > 0\} \tag{D.5}$$

where $\mathbb{1}$ is the indicator function. This is 0-1 loss (Hastie, Tibshirani, & Wainwright, 2015). Sometimes the margin interpretation is used where the log of the conditional probability $\pi_{i,s}$ is used with variables in $\{-1, 1\}$ (see Hastie, Tibshirani, & Friedman, 2001) . Let $z = 2y - 1$ such that for $x \in \{0, 1\}$ we obtain $z \in \{-1, 1\}$. The loss ψ (pseudo log-likelihood) in (D.4) can then be rewritten as

$$\psi(z_i, \mu_i) = \log(1 + \exp(-z_i\mu_i)). \tag{D.6}$$

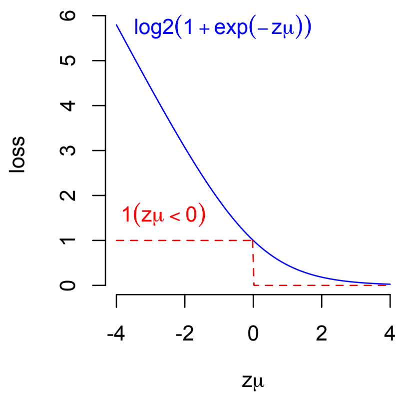


Figure D.1: Misclassification loss $\mathcal{C}$ in (D.7) and ψ in (D.6) as a function of the margin $x\mu = (2y-1)\mu$.

Often the logarithm with base 2 is chosen since then $\psi(z_i, 0) = 1$. The classification translates to

$$\mathcal{C}(z_i) = \mathbb{1}\{z_i\mu_i > 0\}. \tag{D.7}$$

Logistic loss ψ in (D.6) is an upper bound to $\mathcal{C}$ in (D.7), and is 1 at the value of the margin $z_i\mu_i = 0$, as shown in Figure D.1. Here we use logistic loss ψ as defined in (D.4) because it is more common. This function is strictly monotone decreasing.

It follows immediately from monotonicity that $\psi(z_i, \mu) > \psi(z_i, \mu^*)$ if $\mu < \mu^*$. Of course, we have the same for the 0-1 loss: $\mathbb{1}\{z_i\mu_i > 0\} \geq \mathbb{1}\{z_i\mu_i^* > 0\}$ if $\mu < \mu^*$. If the average degree of each node were subtracted from the Hamiltonian μ_s, then we obtain the Ising model without an external field. If we have

$$m_s = -\frac{1}{2}\sum_{t:(s,t)\in E} A_{st} \tag{D.8}$$

then we see that

$$\mu_s = -\frac{1}{2}\sum_{t:(s,t)\in E} A_{st} + \sum_{t:(s,t)\in E} A_{st}x_s x_t = \sum_{t:(s,t)\in E} A_{st}(x_s x_t - 1/2) \tag{D.9}$$

Switching to the labeling $z = 2y-1 \in \{-1, 1\}$, we obtain that the average is 0 (implying $m_s = 0$), and so the Hamiltonian consists only of the interactions

$\sum_{t:(s,t)\in E} A_{st} z_s z_t$. And so if $A_{st} \geq 0$ for all $s,t \in V$ then $\mu_s^* > \mu_s$ iff $A_{st}^* > A_{st}$, i.e., the Hamiltonian is larger if and only of the connectivity is larger. As seen above, this leads immediately to the monotonicity above.

D.1.3 Closeness and Correlations

Let $d_{st}(r) = \min\{1/r_{si} + 1/r_{ij} + \cdots + 1/r_{kt} : \forall i,j,k \in V\backslash\{s,t\}\}$ be the shortest distance in terms of Dijkstra's algorithm (Wallis, 2007), where r_{ij} is the weight, in our case a (polychromic) correlation that are all positive. Then closeness is defined as

$$c_s(r) = \left(\sum_{t \in V\backslash\{s\}} d_{st}(r) \right)^{-1} \tag{D.10}$$

The intuition is here that a node with high closeness will have connections or paths to other nodes with high correlations (weights). To see the intuition, consider node s being connected only to node t with correlation r_{st}. Then $c_s(r) = r_{st}$; if this correlation is high, then so is the closeness of this node. If there is more than one connection, we see that the shortest path $d_{st}(r)$ is low if all correlations r_{ij} are high (close to 1), implying that closeness $c_s(r)$ is high.

Suppose we have two sets weights R_1 and R_2, inducing two graphs G_1 and G_2 with the same nodes and edge sets but with different weights. We pick a path between nodes s and t, denoted by $P_{st} = \{(x_0 = s, x_1), (x_1, x_2), \ldots, (x_{k-1}, x_k = t)\}$ of length k. Suppose that for this path we have

$$\sum_{i=0}^{k} r_{1,i-1,i} \geq \sum_{i=0}^{k} r_{2,i-1,i} \tag{D.11}$$

Then it follows that

$$\sum_{i=0}^{k} \frac{1}{r_{1,i-1,i}} \leq \sum_{i=0}^{k} \frac{1}{r_{2,i-1,i}} \tag{D.12}$$

In other words, the higher the correlations the higher the closeness. This does not imply that any randomly drawn node connected to a node with high closeness will have a high correlation, only that on average the correlations will be higher if they are connected to a node with high closeness than if they are connected to a node with low closeness.

D.2 Supplementary Note 2: Alternative Analysis on Non-Voters

To investigate whether our results are robust to the inclusion of non-voters, we performed the same analyses but now including non-voters and labelling them as voters *against* the focal candidates. The results of this analysis mirrored the results reported in the Results section: The correlation between connectivity and average impact remained high and significant (see Figure D.2a). The same holds for the correlation between centrality and impact (see Figure D.2b). The predicted impact remained very close to the actual impact (deviation median=0.06, deviation interquartile range=0.02-0.09) and outperformed both using the mean of all attitude elements (deviation median=0.10, deviation interquartile range=0.05-0.18, Wilcoxon-matched pairs test: V=4006, P<0.001, CLES=69.5%) and using the means of the specific attitude elements (deviation median=0.08, deviation interquartile range=0.04-0.15, Wilcoxon-matched pairs test: V=5670, P<0.001, CLES=64.7%).

We also tested another prediction from our model regarding differences between voters and non-voters: That voters are expected to have a more densely connected network than non-voters. As can be seen in Figure D.2c, attitude networks of voters were much more highly connected (mean=2.03, s.d.=0.43) than attitude networks of non-voters (mean=2.38, s.d.=0.34, Student's t-test: T=2.86, P<0.001, Cohen's D=0.91).

D.3 Supplementary Note 3: Alternative Analysis on Independents

As our analysis are correlational, it is important to exclude the possibility that third variables affected the relations tested in this paper. The most likely variable to be such a confound is party identification. It is, for example, easy to imagine that party identification might affect the connectivity of attitude networks, the valence of the attitude, and for whom a person votes. We therefore reran our analyses including only participants, who do not identify with any political party. The results of this analysis mirrored the results reported in the Results section: The correlation between connectivity and average impact remained high and significant (see Figure D.3a). The same holds for the correlation between centrality and impact (see Figure D.3b). The predicted impact remained very close to the actual impact (deviation median=0.06, deviation interquartile range=0.03-0.11) and outperformed both using the mean of all attitude el-

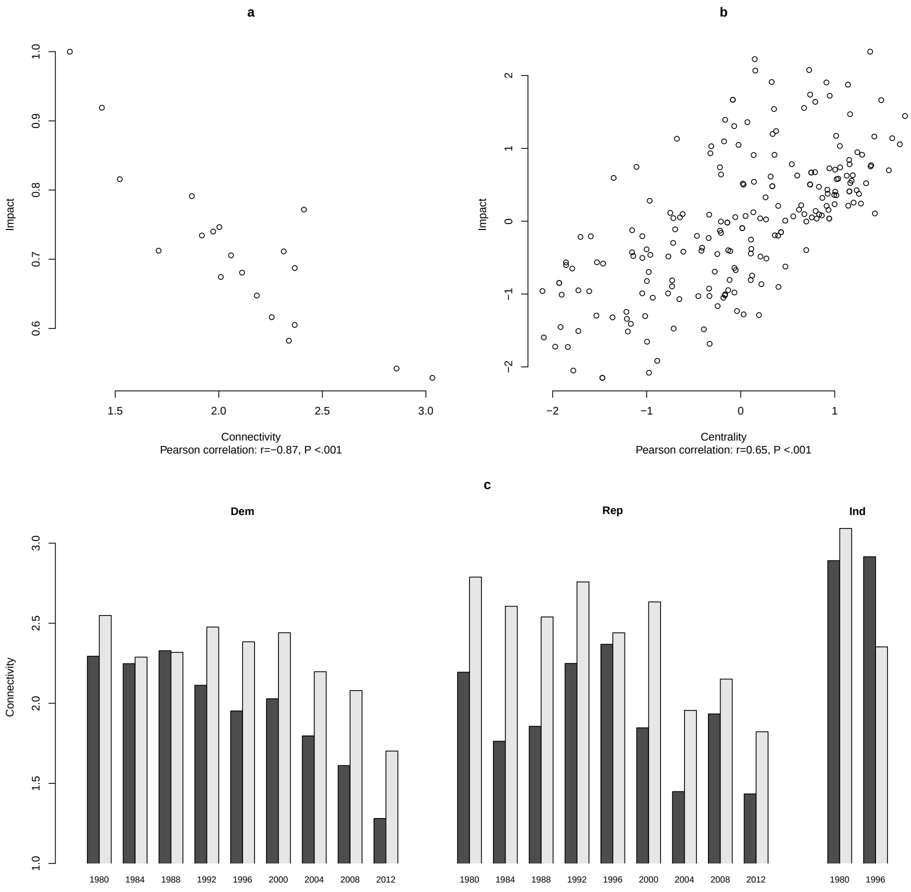


Figure D.2: Analyses including non-voters. (**a**) Relation between connectivity and global impact. (**b**) Relation between centrality and specific impact. (**c**) Comparison between connectivity of voters (dark grey bars) and non-voters (light grey bars). Dem: Democratic candidates; Rep: Republican candidates, Ind: Independent candidates.

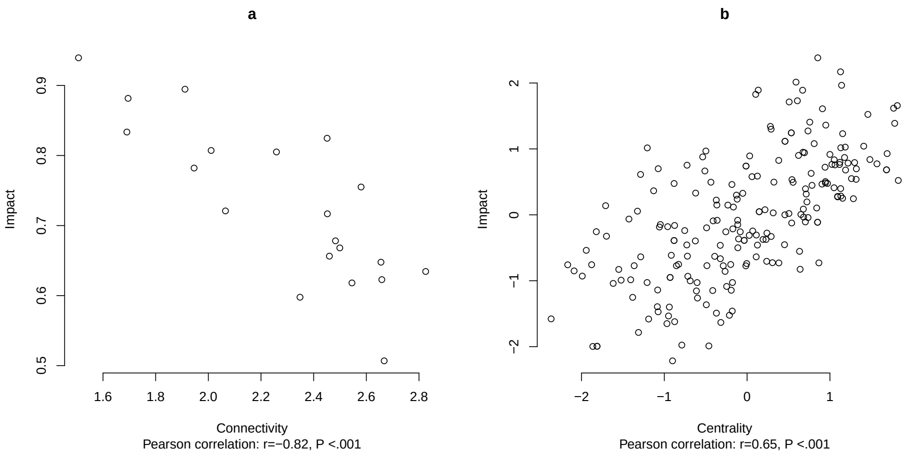


Figure D.3: Analyses including only independents (**a**) Relation between connectivity and global impact. (**b**) Relation between centrality and specific impact.

ements (deviation median=0.11, deviation interquartile range=0.05-0.17, Wilcoxon-matched pairs test: V=4152, P<0.001, CLES=66.2%) and using the means of the specific attitude elements (deviation median=0.09, deviation interquartile range=0.04-0.16), Wilcoxon-matched pairs test: V=5703, P<0.001, CLES=62.7%).

D.4 Supplementary Note 4: Alternative Analysis on Missing Values

To investigate whether our results are robust to imputation of missing values, we reran our analyses with imputing missing values using Predictive Mean Matching (Little, 1988; van Buuren & Groothuis-Oudshoorn, 2011). The results of this analysis mirrored the results reported in the Results section: The correlation between connectivity and average impact remained high and significant (see Figure D.4a). The same holds for the correlation between centrality and impact (see Figure D.4b). The predicted impact remained very close to the actual impact (deviation median=0.06, deviation interquartile range=0.03-0.10) and outperformed both using the mean of all attitude elements (deviation median=0.12, deviation interquartile range=0.05-0.19, Wilcoxon-matched pairs test: V=3724, P<0.001, CLES=69.3%) and using the means of the specific attitude elements (deviation median=0.10, deviation interquartile range=0.04-0.18, Wilcoxon-matched pairs test: V=5337, P<0.001, CLES=65.5%).

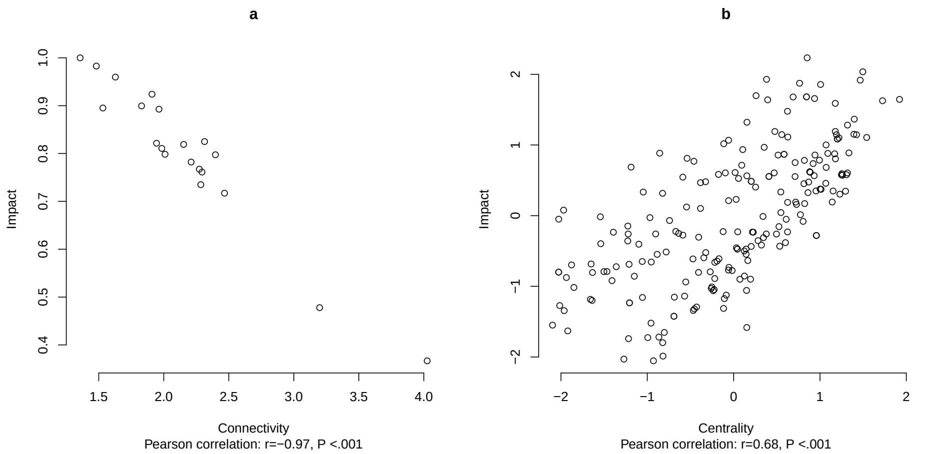


Figure D.4: Analyses with imputed missing values (**a**) Relation between connectivity and global impact. (**b**) Relation between centrality and specific impact.

D.5 Supplementary Note 5: Alternative Analysis on Networks Based on Different Numbers of Attitude Elements

To investigate whether our results are robust to our choice of attitude elements, we reran our analyses based on all available attitude elements (note that in this case the forecast analyses are not possible because for these analyses the same number of attitude elements for each election is necessary) and on the seven attitude elements that were assessed at each election. The results of these analyses mirrored the results reported in the Results section: For the analysis including all attitude elements, the correlation between connectivity and average impact remained high and significant (see Figure D.5a). The same holds for the correlation between centrality and impact (see Figure D.5b).

For the analysis including the seven attitude elements that were assessed at each election, the correlation between connectivity and average impact remained high and significant (see Figure D.5c). The same holds for the correlation between centrality and impact (see Figure D.5d). The predicted impact remained very close to the actual impact (deviation median=0.06, deviation interquartile range=0.03-0.09) and outperformed both using the mean of all attitude elements (deviation median=0.11, deviation interquartile range=0.06-0.18, Wilcoxon-matched pairs test: V=1505, P<0.001, CLES=68.7%) and using the means of the

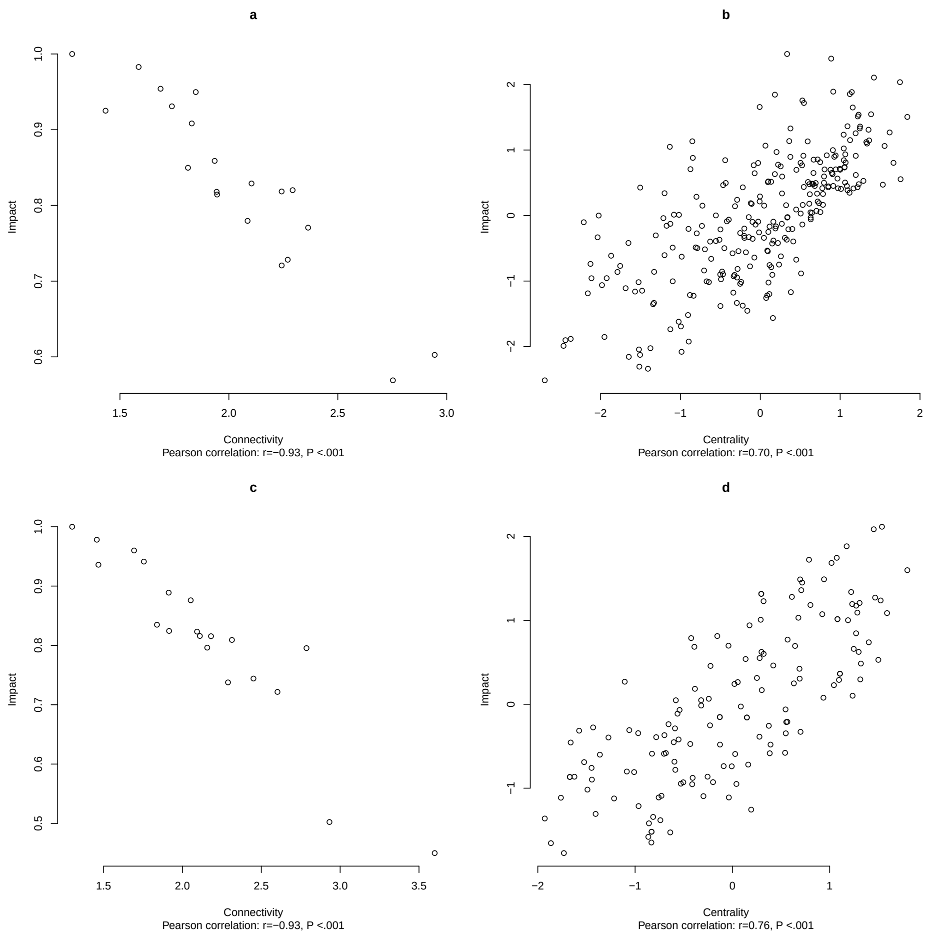


Figure D.5: Analyses on different numbers of attitude elements (**a**) Relation between connectivity and global impact for the analysis including all attitude elements. (**b**) Relation between centrality and specific impact for the analysis including all attitude elements. (**c**) Relation between connectivity and global impact for the analysis including seven attitude elements. (**d**) Relation between centrality and specific impact for the analysis including seven attitude elements.

specific attitude elements (deviation median=0.09, deviation interquartile range=0.05-0.16), Wilcoxon-matched pairs test: V=2341, P<0.001, CLES=65.2%).

D.6 Supplementary Note 6: Alternative Analysis on Latent Variable Models

One might argue that the results reported in this article can also be expected when attitudes are conceptualized as latent variables and the responses on attitude elements are treated as indicators of the latent attitude. From this perspective, high (low) centrality of attitude elements would indicate high (low) factor loadings on the latent variable attitude and high (low) connectivity would indicate high (low) average factor loadings. In a purely statistical sense, this objection would be correct as factor loadings also reflect how much information a given attitude element holds on all other attitude elements. However, in our view the latent variable framework does not provide a sensible alternative for a data-generating model of the hypotheses put forward here. For such a model, one would have to assume that the latent variable attitude acts as common cause of the attitude elements. This assumption, however, is at odds with several key concepts in the attitude literature (Dalege et al., 2016), such as cognitive consistency (Monroe & Read, 2008), ambivalence (Thompson et al., 1995), and the idea that attitudes are formed by attitude elements (Fazio, 1995; Zanna & Rempel, 1988).

To further rule out that the latent variable framework provides an alternative explanation of our results, we investigated the fit of latent variable models on the data reported in this article. For each attitude toward each candidate at each election, we fitted two latent variable models. First, we fitted a one-factor model with all attitude elements loading on this single factor representing a latent attitude. Second, we fitted a hierarchical factor model with three or four first-order factors and one second-order factor representing a latent attitude. We fitted the hierarchical factor model because earlier research indicated that beliefs and feelings form different factors (Breckler, 1984) and that negative and positive attitude elements form different factors (van den Berg et al., 2005). In most of the data sets used here, no negative beliefs were assessed. For these data sets, we fitted a hierarchical factor model with beliefs, negative feelings, and positive feelings loading on different first-order factors, respectively. For the data sets in which negative beliefs were assessed, we fitted a hierarchical factor model with negative beliefs, positive beliefs, negative feelings,

Table D.1: Results of the Simulations for Each Network-Generating Algorithm and Edge Weight Distribution.

	Preferential attachment	Small-world	Random graph
	Connectivity/impact correlations		
Normal distribution	mean r=-0.91	mean r=-0.91	mean r=-0.90
	s.d. r=0.07	s.d. r=0.05	s.d. r=0.08
Power-law distribution	mean r=-0.92	mean r=-0.91	mean r=-0.91
	s.d. r=0.05	s.d. r=0.05	s.d. r=0.04
Uniform distribution	mean r=-0.92	mean r=-0.89	mean r=-0.90
	s.d. r=0.05	s.d. r=0.08	s.d. r=0.08
	Centrality/impact correlations		
Normal distribution	mean r=0.72	mean r=0.51	mean r=0.57
	s.d. r=0.18	s.d. r=0.33	s.d. r=0.27
Power-law distribution	mean r=0.70	mean r=0.46	mean r=0.60
	s.d. r=0.19	s.d. r=0.34	s.d. r=0.23
Uniform distribution	mean r=0.68	mean r=0.49	mean r=0.60
	s.d. r=0.24	s.d. r=0.29	s.d. r=0.25

and positive feelings loading on different first-order factors, respectively. As can be seen in Table D.3, both the one-factor models and the hierarchical models fitted poorly. The latent variable framework thus appears to be an unlikely alternative explanation of our results.

This discussion on whether our results can also be explained by the latent variable framework is somewhat reminiscent of the discussion regarding the idea that instability of attitudinal responses is indicative of individuals holding nonattitudes (Converse, 1970). Several critiques of this idea pointed out that when measurement error is accounted for, individuals, who seemingly hold nonattitudes, show stable attitudes (Achen, 1975; Ansolabehere, Rodden, & Snyder, 2008; Judd, Krosnick, & Milburn, 1980). A similar critique might apply to our findings. It is our view, however, that two findings speak against this critique. First, if we assume that the intercorrelations of attitude elements are determined only (or foremost) by measurement error, then the factor models we fitted should show good fit. This was clearly not the case. Second, systematic variation of intercorrelations would not be expected from the measurement error perspective. Thus, our finding that connectivity of attitudes correlates almost perfectly with the attitude's impact on behaviour would not be expected from the measurement error perspective.

Table D.2: Number of participants per election and number of participants with missing values.

Election	N complete sample	N non-voters	N with missing values (Democratic candidates)	N with missing values (Republican candidates)
1980	1,614	411	338	396
1984	2,257	539	583	435
1988	2,040	545	470	477
1992	2,485	562	567	358
1996	1,714*	374	239	318
2000	1,807	376	440	493
2004	1,212	231	282	195
2008	2,322	509	368	390
2012	5,914	1,141	557	644

Note. *Of these 1,714 participants, 1,316 participants also participated during the election of 1992.

Table D.3: Fit measures of the latent variable models.

Candidate	One-factor model	Hierarchical model
Carter 1980	$\chi(35) = 1030.53$, $CFI = 0.80$, $RMSEA = 0.18$	$\chi(30) = 528.58$, $CFI = 0.90$, $RMSEA = 0.14$
Reagan 1980	$\chi(35) = 411.40$, $CFI = 0.91$, $RMSEA = 0.12$	$\chi(30) = 198.74$, $CFI = 0.96$, $RMSEA = 0.08$*
Anderson 1980	$\chi(35) = 655.19$, $CFI = 0.75$, $RMSEA = 0.18$	$\chi(30) = 290.38$, $CFI = 0.90$, $RMSEA = 0.12$*
Mondale 1984	$\chi(35) = 1410.69$, $CFI = 0.79$, $RMSEA = 0.19$	$\chi(31) = 859.75$, $CFI = 0.87$, $RMSEA = 0.15$
Reagan 1984	$\chi(35) = 1730.83$, $CFI = 0.83$, $RMSEA = 0.19$	$\chi(31) = 1194.73$, $CFI = 0.88$, $RMSEA = 0.17$*
Dukakis 1988	$\chi(35) = 1166.68$, $CFI = 0.81$, $RMSEA = 0.18$	$\chi(31) = 683.11$, $CFI = 0.89$, $RMSEA = 0.14$
Bush 1988	$\chi(35) = 940.98$, $CFI = 0.87$, $RMSEA = 0.16$	$\chi(31) = 593.40$, $CFI = 0.92$, $RMSEA = 0.13$*
Clinton 1992	$\chi(35) = 1536.91$, $CFI = 0.83$, $RMSEA = 0.18$	$\chi(31) = 995.36$, $CFI = 0.89$, $RMSEA = 0.15$*
Bush 1992	$\chi(35) = 1304.31$, $CFI = 0.85$, $RMSEA = 0.15$	$\chi(31) = 840.63$, $CFI = 0.91$, $RMSEA = 0.13$*
Clinton 1996	$\chi(35) = 1492.00$, $CFI = 0.83$, $RMSEA = 0.19$	$\chi(31) = 908.92$, $CFI = 0.90$, $RMSEA = 0.16$*
Dole 1996	$\chi(35) = 1312.91$, $CFI = 0.80$, $RMSEA = 0.19$	$\chi(31) = 842.52$, $CFI = 0.87$, $RMSEA = 0.16$
Perot 1996	$\chi(35) = 428.94$, $CFI = 0.78$, $RMSEA = 0.18$	Fit measures could not be computed*
Gore 2000	$\chi(35) = 1401.65$, $CFI = 0.79$, $RMSEA = 0.19$	$\chi(31) = 837.05$, $CFI = 0.88$, $RMSEA = 0.16$
Bush 2000	$\chi(35) = 1080.12$, $CFI = 0.84$, $RMSEA = 0.18$	$\chi(31) = 685.86$, $CFI = 0.90$, $RMSEA = 0.15$
Kerry 2004	$\chi(35) = 771.78$, $CFI = 0.86$, $RMSEA = 0.17$	$\chi(31) = 458.36$, $CFI = 0.92$, $RMSEA = 0.14$*
Bush 2004	$\chi(35) = 1100.80$, $CFI = 0.87$, $RMSEA = 0.20$	$\chi(31) = 661.61$, $CFI = 0.92$, $RMSEA = 0.16$*
Obama 2008	$\chi(35) = 1373.76$, $CFI = 0.90$, $RMSEA = 0.16$	$\chi(31) = 642.16$, $CFI = 0.95$, $RMSEA = 0.12$*
McCain 2008	$\chi(35) = 1553.26$, $CFI = 0.86$, $RMSEA = 0.18$	$\chi(31) = 863.92$, $CFI = 0.92$, $RMSEA = 0.14$
Obama 2012	$\chi(35) = 5863.21$, $CFI = 0.91$, $RMSEA = 0.20$	$\chi(31) = 2428.10$, $CFI = 0.96$, $RMSEA = 0.14$
Romney 2012	$\chi(35) = 5781.98$, $CFI = 0.88$, $RMSEA = 0.20$	$\chi(31) = 3021.35$, $CFI = 0.94$, $RMSEA = 0.15$

Note. *The covariance matrices of the latent variables of these models were not positive definitive. This indicates poor fit to the data.

Appendix E

Supplementary Information on Simulation 1b of Chapter 6

Table E.1: Probability of Assignment of Subject with Given Threshold to a Given Group in Simulation 1b.

	Thresholds									
Group	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
#1	.19	.17	.15	.13	.11	.09	.07	.05	.03	.01
#2	.17	.15	.14	.12	.11	.09	.08	.06	.05	.03
#3	.15	.14	.13	.12	.11	.09	.08	.07	.06	.05
#4	.13	.12	.12	.11	.10	.10	.09	.08	.08	.07
#5	.11	.11	.11	.10	.10	.10	.10	.09	.09	.09
#6	.09	.09	.10	.10	.10	.10	.10	.11	.11	.11
#7	.07	.08	.08	.09	.10	.10	.11	.12	.12	.13
#8	.05	.06	.07	.08	.10	.11	.12	.13	.14	.15
#9	.03	.05	.06	.08	.09	.11	.12	.14	.15	.17
#10	.01	.03	.05	.07	.09	.11	.13	.15	.17	.19

Appendix F

Contributions to Chapters

Chapter 2

Dalege, J., Borsboom, D., van Harreveld, F., van der Berg, H., Conner, M., & van der Maas, H. L. J. (2016). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychological Review, 123*, 2-22.

J. Dalege and H. L. J. van der Maas developed the main idea for this manuscript; D. Borsboom and F. van Harreveld contributed to the further development of the idea; J. Dalege wrote the manuscript and analyzed the data; D. Borsboom, F. van Harreveld, H. van der Berg, M. Conner, and H. L. J. van der Maas provided valuable comments and revisions.

Chapter 3

Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2017). Network analysis on attitudes: A brief tutorial. *Social Psychological and Personality Science, 8*, 528-527.

J. Dalege developed the concept for this manuscript, anlayzed the data, and wrote the manuscript; D. Borsboom, F. van Harreveld, and H. L. J. van der Maas provided valuable comments and revisions.

Chapter 4

Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2018). A network perspective on attitude strength: Testing the connectivity hypothesis. *Social Psychological and Personality Science*. Advance online publication.

J. Dalege developed the study concept; J. Dalege, D. Borsboom, F. van Harreveld, and H. L. J. van der Maas contributed to the study design; J. Dalege performed the data analysis and interpretation; J. Dalege wrote the manuscript; D. Borsboom, F. van Harreveld, and H. L. J. van der Maas provided valuable comments and revisions.

Chapter 5

Dalege, J., Borsboom, D., van Harreveld, F., Waldorp, L. J. & van der Maas, H. L. J. (2017). Network structure explains the impact of attitudes on voting decisions. *Scientific Reports*, *7*, 4909.

J. Dalege developed the study concept; J. Dalege, D. Borsboom, F. van Harreveld, and H. L. J. van der Maas contributed to the study design; J. Dalege performed the data analysis and interpretation; J. Dalege wrote the manuscript; D. Borsboom, F. van Harreveld, and H. L. J. van der Maas provided valuable comments and revisions; Lourens J. W. provided the analytical solutions of the hypotheses.

Chapter 6

Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2018). The Attitudinal Entropy (AE) framework as a general theory of individual attitudes. *Psychological Inquiry*, *29*, 175-193.

J. Dalege developed the main idea for this manuscript; D. Borsboom, F. van Harreveld, and H. L. J. van der Maas contributed to the further development of the idea; J. Dalege wrote the manuscript and performed the simulations; D. Borsboom, F. van Harreveld, and H. L. J. van der Maas provided valuable comments and revisions.

Chapter 7

Dalege, J., Borsboom, D., van Harreveld, F., Lunansky, G., & van der Maas, H. L. J. (2018). The Attitudinal Entropy (AE) framework: Clarifications, extensions, and future directions. *Psychological Inquiry*, *29*, 218-228.

J. Dalege developed the main ideas for this manuscript; D. Borsboom, F. van Harreveld, and H. L. J. van der Maas contributed to the further development of the ideas; J. Dalege wrote the manuscript and performed the simulation; J. Dalege and G. Lunansky developed the app; D. Borsboom, F. van Harreveld, and H. L. J. van der Maas provided valuable comments and revisions.

Chapter 8

Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2018). *The Learning Ising Model of Attitude (LIMA): Entropy reduction by Hebbian learning*. Manuscript in preparation.

J. Dalege and H. L. J. van der Maas developed the main idea for this manuscript; D. Borsboom and F. van Harreveld contributed to the further development of the idea; J. Dalege and H. L. J. van der Maas performed the simulations; J. Dalege wrote the manuscript; D. Borsboom, F. van Harreveld, and H. L. J. van der Maas provided valuable comments and revisions.

Appendix G

Summary

G.1 A Formal Approach to Attitude

Attitudes, the liking or disliking of an object, are formed through several different processes (e.g., Fazio, 1995, 2007; Zanna & Rempel, 1988), range from being stable and impactful to being fluctuating and inconsequential (e.g., Krosnick & Petty, 1995), and are related to several core constructs in social psychology – in short, attitudes are highly complex. Given the complexity of attitude it might seem that it is in vein to try to develop formal theories on attitudes. Yet, recent developments in psychology specifically and complexity science more generally have shown that complex behavior can be the result of relatively simple processes interacting with each other. In this dissertation, I develop and test several formalisms on attitude relying on few fundamental principles.

The approach formulated in this dissertation rests on two fundamental assumptions. First, attitude elements (i.e., beliefs, feelings, and behaviors vis-á-vis an attitude object) directly interact with each other (e.g., judging snakes as dangerous leads to feeling afraid of snakes) and these interactions can be modeled by representing them in a network structure. Second, interactions between attitude elements become more pronounced as more attention and thought is directed at the attitude object, implying that attitudes are unstable and inconsistent (high in entropy) when one directs little attention to the attitude object, while attitudes become more stable and consistent (low in entropy) when one thinks frequently about the attitude object.

Part I introduces the Causal Attitude Network (CAN) model, which represents a measurement model of attitude based on the assumptions that attitude elements form a network of direct causal interactions. Chapter 2 provides a detailed description of the CAN model, which uses the Ising (1925) model to simulate the dynamics of attitudes in an idealized fashion. I discuss the CAN model's assumption that the structure of attitude networks conforms to a small-world structure in which attitude elements similar to each other form tight clusters. Attitude strength is formalized as network connectivity in the CAN model and I show that this formalism provides a parsimonious explanation of the differences between strong and weak attitudes.

Chapter 3 provides a tutorial on how to apply the CAN model to empirical data and how to perform network analysis on attitude data. First, I show how one can estimate attitude networks and use data on attitudes toward Barack Obama during the election of 2012 as an illustration. Second, I discuss how one can calculate standard network measures, such as community structure, centrality, and connectivity. Third, I show how one can simulate data from an estimated attitude network and illustrate consequences of varying connectivity and of targeting central versus peripheral attitude elements.

Part II provides empirical tests of the CAN model. Chapter 4 tests the central assumption of the CAN model that highly connected attitude networks correspond to strong attitudes. For this, I use data from the American National Election Studies (ANES) from 1980 to 2012 on attitudes toward presidential candidates. First, I estimate attitude networks of groups differing in political interest and show that political interest and network connectivity are highly related. I then test whether network connectivity is related to two defining features of attitude strength – attitude stability and impact on behavior. I show that network connectivity is strongly positively related to attitude stability and the attitude's impact on behavior, providing support for the CAN model's hypothesis that highly connected attitude networks conform to strong attitudes.

Chapter 5 illustrates the CAN model's applied value by testing whether network structure can be used to forecast impact of specific attitude elements and impact of the global attitude on voting decisions. Using simulations, I first show that the CAN model predicts that highly connected attitude networks have the strongest impact on behavior globally and that in any attitude network central attitude elements have the highest impact on behavior. I test these hypotheses on data from the ANES from 1980 to 2012 and find strong support for these hypotheses. Additionally, I show that the CAN model allows one to forecast with high accuracy the specific impact of attitude elements on voting decisions.

Part III introduces the Attitudinal Entropy (AE) framework as a general theory of attitudes. Chapter 6 discusses the main principles of the AE framework, which builds on the CAN model, and connects the basic principles of statistical mechanics to attitude research by means of analogical modeling. The AE framework rests on three fundamental principles. First, attitude inconsistency and instability represent two related indicators of attitudinal entropy, which implies that attitudes gravitate towards inconsistent and unstable states. Second, energy of attitudinal representations serves as a local mechanism to reduce the global entropy of attitudes. Third, attention and thought have a similar effect as (inverse) temperature has in thermodynamic systems – thought and attention reduce attitudinal entropy. I show that several well-established effects in the attitude literature, such as the mere thought effect (Tesser, 1978) and heuristic versus systematic processing of arguments under low versus high involvement (e.g., Petty & Cacioppo, 1986), follow from the AE framework. Additionally, I discuss the AE framework's implications for ambivalence and cognitive dissonance.

Chapter 7 provides a reply to several commentaries on Chapter 6. In this reply, I first discuss the aim of the AE framework and provide some clarifications of its main principles. Second, I discuss the AE framework's implications for dual-process models and evaluate whether the AE framework provides novel predictions. Third, I apply the AE framework to measurement of attitudes and to the interpersonal realm. Finally, I provide an online app, in which the basics of the AE framework are implemented, to make the formalisms of the AE framework more accessible.

Chapter 8 introduces the Learning Ising Model of Attitude (LIMA), which adds Hebbian learning to the AE framework. The LIMA thereby provides an explanation for how attitude networks, which are able to reduce entropy, develop. In several simulations I illustrate the dynamics of the LIMA. First, I show that Hebbian learning leads to networks, which efficiently reduce entropy. Second, I model feedback between instability of attitudes and the attention directed at the attitude object, implement learning of dispositions of attitude elements, and provide an illustration of attitude change. I discuss the LIMA's contribution to the unification of research on attitudes, its relation to Hopfield network models and Boltzmann machines, and potential avenues for future research.

I end by discussing (1) how the different formalisms developed in this dissertation relate to each other, (2) how the formalisms might be applied to intra- and interpersonal interactions between attitudes, (3) on what time scales the formalisms operate, and (4) the possibility of estimating individual networks. I then place the formalisms developed in this dissertation in a broader perspective and discuss their implications for the nature of attitudes. I argue that the essence of attitudes is a structural one and that successful theories on attitudes can only illuminate the structure of attitudes but not the properties of the constituents making up this structure. I then discuss measurement from the perspective of the approach of this dissertation, which might lead to novel definitions of measurement error and measurement bias. Additionally, I connect this dissertation's approach to coarse-graining (i.e., the smoothing out of irrelevant information to arrive at an effective theory) and by this provide a possible avenue for a novel perspective on mental causation. Finally, I discuss how the approach of this dissertation relates to structural realism in philosophy of science. I conclude that the principles developed in this dissertation not only provide the building blocks for a theory of attitude, but can also serve as a blueprint for the dynamics of mental representations in general.

Appendix H

Nederlandse Samenvatting

H.1 Een Formele Benadering van Attitude

Attitudes, de voorkeur of afkeur ten opzichte van een object, worden gevormd door velen verschillende processen (bijvoorbeeld Fazio, 1995, 2007; Zanna & Rempel, 1988), kunnen stabiel en invloedrijk of fluctuerend en irrelevant zijn (bijvoorbeeld Krosnick & Petty, 1995), en zijn gerelateerd aan velen fundamentele concepten binnen de sociale psychologie – in het kort, attitudes zijn erg complex. Door deze complexiteit zou men geneigd kunnen zijn te concluderen dat het hopeloos is formele theorieën van attitudes te ontwikkelen. Echter, recente ontwikkelingen in de psychologie en de complexiteit wetenschap hebben laten zien dat complex gedrag de uitkomst kan zijn van relatief simpele processen die op elkaar inwerken. In dit proefschrift ontwikkel en toets ik een formele benadering van attitude die op een klein aantal fundamentele principes berust.

De benadering van dit proefschrift berust op twee fundamentele assumpties. Ten eerste, attitude elementen (meningen, gevoelens, en gedragingen ten opzichte van een attitude object) werken direct op elkaar in (bijvoorbeeld leidt dat je denkt dat slangen gevaarlijk zijn tot een gevoel van angst als je een slang ziet) en deze wederzijdse inwerkingen kunnen als netwerken worden gemodelleerd. Ten tweede, wederzijdse inwerkingen tussen attitude elementen worden versterkt wanneer iemand zijn attentie op het attitude object richt en nadenkt over het attitude object. Hierdoor zijn attitudes instabiel en inconsistent (hoog in entropie) wanneer iemand weinig attentie op het attitude object richt, terwijl attitudes stabieler en consistenter (laag in entropie) worden wanneer iemand vaak over het attitude object nadenkt.

Deel I introduceert het *Causal Attitude Network* (CAN) model. Dit model representeert een meetmodel van attitude dat op de assumptie berust dat attitude elementen een netwerk van wederzijdse causale inwerkingen vormen. Hoofdstuk 2 beschrijft de details van het CAN model dat het Ising (1925) model gebruikt om attitude dynamieken op een geïdealiseerde manier te simuleren. Ik bespreek de assumptie van het CAN model dat de structuur van attitude netwerken zich voegt in een *small-world* structuur waarin attitude elementen die soortgelijk zijn clusters vormen. Attitude sterkte wordt in het CAN model als netwerk connectiviteit geformaliseerd en ik laat zien dat dit formalisme een spaarzame verklaring biedt voor de verschillen tussen sterke en zwakke attitudes.

Hoofdstuk 3 biedt een aanleiding hoe men het CAN model op empirische data toe kan passen en hoe men netwerk analyses op attitude data kan toepassen. Ten eerste laat ik zien hoe men attitude netwerken kan schatten. Dit wordt geïllustreerd aan hand van data van attitudes ten opzichte van Barack Obama tijdens de Amerikaanse verkiezingen van 2012. Ten tweede bespreek ik hoe men netwerk maten, zoals *community* structuren, centraliteit, en connectiviteit, kan berekenen. Ten deerde laat ik zien hoe men data vanuit een attitude netwerk kan simuleren en hierbij illustreer ik de consequenties van variaties in netwerk connectiviteit en het beïnvloeden van centrale versus perifere attitude elementen.

Deel II biedt empirische toetsen van het CAN model. Hoofdstuk 4 toetst de centrale hypothese van het CAN model dat attitude netwerken met hoge connectiviteit overeenkomen met sterke attitudes. Voor deze toets gebruik ik data van de *American National Election Studies* (ANES) van 1980 tot en met 2012. Deze data focust op attitudes ten opzichte van kandidaten voor het presidentschap. Ten eerste schat ik netwerken van groepen die in hun politieke interesse verschillen en laat ik zien dat politieke interesse en netwerk connectiviteit sterk positief zijn gerelateerd. Dan toets ik of netwerk connectiviteit is gerelateerd aan twee centrale eigenschappen van attitude sterkte – de stabiliteit van attitudes en het invloed van attitudes op gedrag. Ik laat zien dat netwerk connectiviteit sterk positief samenhangt met de stabiliteit van attitudes en het invloed van attitudes op gedrag. Deze bevinding ondersteund de hypothese van het CAN model dat attitude netwerken met hoge connectiviteit overeenkomen met sterke attitudes.

Hoofdstuk 5 illustreert de toegepaste waarde van het CAN model door te toesten of de netwerk structuur van attitudes kan worden gebruikt om te voorspellen welke specifieke attitude elementen de sterkste invloed op verkiezingsgedrag hebben en te voorspellen of de globale attitude invloed op verkiezingsgedrag heeft. Met behulp van simulaties laat ik eerst zien dat het CAN model voorspelt dat attitude netwerken met hoge connectiviteit globaal de sterkste invloed op gedrag hebben en dat binnen attitude netwerken de meest centrale attitude elementen de sterkste invloed op gedrag hebben. Ik toets deze hypotheses aan de hand van data van de

ANES van 1980 tot en met 2012. Deze toetsen laten sterke ondersteuning voor de hypotheses van het CAN model zien. Daarnaast laat ik zien dat men met behulp van het CAN model de specifieke invloed van attitude elementen op het verkiezingsgedrag met hoge nauwkeurigheid kan voorspellen.

Deel III introduceert een algemene theorie van attitude: het *Attitudinal Entropy (AE) framework*. Hoofdstuk 6 bespreekt de fundamentele principes van het *AE framework* dat voortbouwt op het CAN model. Door analoog modelleren verbindt het *AE framework* de fundamentele principes van de statistische mechanica met het onderzoek naar attitudes. Het *AE framework* baseert op drie fundamentele principes. Ten eerste, inconsistentie en instabiliteit van attitudes representeren twee gerelateerd indicatoren van entropie van attitudes. Dit impliceert dat attitudes naar inconsistente en instabiele toestanden graviteren. Ten tweede, energie van attitude representaties dient als een lokaal mechanisme om de globale entropie van attitudes te reduceren. Ten deerde, attentie en nadenken hebben een effect dat is vergelijkbaar met het effect van (inverse) temperatuur in thermodynamische systemen – attentie en nadenken verlagen de entropie van attitudes. Ik laat zien dat velen gevestigde effecten in de attitude literatuur, zoals het *mere thought* effect (bijvoorbeeld Tesser, 1978) en dat betrokkenheid determineert of argumenten heuristisch of systematisch worden verwerkt (bijvoorbeeld Petty & Cacioppo, 1986), uit het *AE framework* kunnen worden afgeleid. Daarnaast bespreek ik de implicaties van het *AE framework* voor ambivalentie en cognitieve dissonantie.

Hoofdstuk 7 biedt een respons op commentaren op Hoofdstuk 6. In dit respons bespreek ik eerst het doel van het *AE framework* en biedt een aantal verduidelijkingen van de fundamentele principes van het *AE framework*. Ten tweede bespreek ik de implicaties van het *AE framework* voor duale procesmodellen en ik evalueer of het *AE framework* nieuwe voorspellingen biedt. Ten deerde pas ik het *AE framework* toe op de meting van attitudes en op interpersoonlijke factoren die invloed op attitudes hebben. Ten slot biedt het hoofdstuk een online app waarin de basics van het *AE framework* zijn geïmplementeerd. Hierdoor worden de formalismen van het *AE framework* toegankelijker gemaakt.

Hoofdstuk 8 introduceert het *Learning Ising Model of Attitude* (LIMA) dat Hebbian leren aan het *AE framework* toevoegt. Het LIMA biedt een verklaring hoe attitude netwerken, die entropie van attitudes kunnen verlagen, ontstaan. In een aantal simulaties illustreer ik de dynamieken van het LIMA. Ten eerste laat ik zien dat Hebbian leren tot attitude netwerken, die effectief entropie verlagen, leidt. Ten tweede modeleer ik feedback tussen instabiliteit van attitudes en het richten van attentie op attitude objecten, implementeer ik het leren van disposities van attitude elementen, en biedt ik een illustratie van attitude verandering binnen het LIMA. Ik bespreek de contributies van het LIMA aan de unificatie van attitude onderzoek, hoe het LIMA relateert aan Hopfield netwerk modellen en Boltzmann machines, en mogelijkheden voor toekomstig onderzoek naar het LIMA.

Ik sluit met het bespreken van (1) hoe de verschillende formalismen die in dit proefschrift werden ontwikkeld tot elkaar relateren, (2) hoe de formalismen wellicht kunnen worden toegepast op intra- en interpersoonlijke inwerkingen tussen attitudes, (3) op welke tijd schalen de verschillende formalismen werken, en (4) de mogelijkheden om individuele attitude netwerken te schatten. Daarna plaats ik de formalismen die in dit proefschrift werden ontwikkelt in een breder kader en ik bespreek de implicaties van deze formalismen voor de natuur van attitudes. I betoog dat de essentie van attitudes van structureel aard is en dat succesvolle theorieën van attitude slechts de structuur van attitudes kunnen verduidelijken maar niet de elementen die deze structuur opmaken. Daarna bespreek ik psychologische meting vanuit de benadering van dit proefschrift en introduceer hierdoor nieuwe definities van meetfout een meetvertekeningen. Daarnaast koppel ik de benadering van dit proefschrift aan *coarse-graining* (het uitfilteren van irrelevante informatie om een effectieve theorie te ontwikkelen) en daardoor biedt ik een nieuw perspectief op mentale oorzakelijkheid. Tot slot bespreek ik hoe de benadering van dit proefschrift relateert aan structureel realisme, een stroming binnen de wetenschapsfilosofie. Ik concludeer dat de principes die in dit proefschrift werden geformuleerd niet alleen de bouwstenen voor een theorie van attitude representeren, maar wellicht ook een blueprint voor de dynamieken van mentale representaties in het algemeen bieden.

Appendix I

Acknowledgments – Dankwoord

Dit proefschrift was niet mogelijk geweest zonder de velen mensen die mij hebben ondersteund. Allereerst wil ik mijn begeleiders voor een bijzonder fijne manier van begeleiding danken. Han, Denny, Frenk, voordat ik mijn dank aan jullie individueel uitspreek wil ik mijn dankbaarheid voor de inspirerende, uitdagende, en plezierige meetings uitspreken. Elke keer na een meeting met jullie voelde ik me een stuk wijzer en gelukkiger dan voor de meeting. Het is onmogelijk om voldoende op prijs te stellen hoe privilegieert ik ben geweest regelmatig met zulke briljante mensen, die ook nog expertise in verschillende gebieden hebben, te mogen discuteren. Ik denk dat we in elke meeting interessante en belangrijke connecties tussen verschillende gebieden hebben gemaakt en deze connecties waren de basis van de interdisciplinaire insteek van dit proefschrift.

Han, we werken nu al seven jaar samen en ik waardeer elke minuut hiervan. Ik kan mij nog goed aan de cursus herinneren waarmee onze samenwerking begon en hoe zeer ik onder de indruk was hoe makkelijk jij connecties kan maken tussen allerlei verschillende vakgebieden. Dit ben ik trouwens nog steeds! Naast dat ik me geen betere wetenschapper dan jou kan voorstellen, ben je ook nog de perfecte begeleider. Ik had altijd het gevoel dat ik veel vrijheid had om mijn ideeën te ontwikkelen, maar ook dat ik altijd op goed feedback van jou kon rekenen. Ik vindt het altijd erg vermakelijk als je mijn kantoor binnenloopt en wij over onze projecten of allerlei andere interessante dingen discuteren. Ik denk ook dat we een diepe connectie hebben hoe we over wetenschappelijke vraagstukken nadenken en ik kijk uit naar al de ideeën die we nog samen zullen ontwikkelen!

Denny, als je bij google het woord genius intypt zou eigenlijk een foto van jou moeten verschijnen. In ieder geval ben je voor mij een incarnatie van de geniale wetenschapper! Ik snap nog steeds niet helemaal hoe je het doet, maar bij elke meeting heb je ideeën die een radicaal nieuw perspectief bieden. Ik was trouwens al onder de indruk van jou tijdens de eerste week van mijn studie toen je een college van het vak Methodenleer gaf. Het enthousiasme waarmee je toen over psychologisch onderzoek hebt gepraten is wellicht de hoofdreden waarom ik in de eerste week van mijn studie heb besloten dat ik psychologisch onderzoeker wilde worden.

Frenk, jouw passie voor attitude onderzoek is een van de grootste redenen dat ik onderzoek naar attitudes ging doen. Ook jij had namelijk al voordat ik aan mij PhD begon een grote invloed op mijn wetenschappelijke carrière. Ik kan me nog goed herinneren hoe een les van jou mij eerst had overtuigd om sociale psychologie als specialisatie te kiezen en daarna heeft jou fantastische collegereeks over attitudes mijn passie voor dit onderwerp gewekt. Je was ins ons groepje diegene met het oog voor details en ik ben erg dankbaar voor jouw kritisch maar constructief feedback op het werk in dit proefschrift. Zonder jou was dit proefschrift veel minder sterk gegrond in de attitude literatuur.

Mark, Juliane, Agneta, Gerben, and Lourens, thank you for reading and evaluating my dissertation. I very much appreciate it to have such an illustrious group of scientists as my dissertation committee!

Juliane, vielen Dank dafür, dass du meine wissenschaftliche Karriere schon früh in meinem Studium gefördert hast und für dein Verständnis, dass ich wieder nach Amsterdam zurückgekehrt bin. Es freut mich sehr, dass du Teil meiner Kommission bist und mich somit auch auf diesem Teil meiner Karriere begleitest.

Marie en Sacha, ik vind het erg leuk dat jullie mijn paranimfen zijn. Jullie hebben allebei een grote rol gespeld tijdens mijn hele PhD!

Marie, ohne die unzähligen Kaffeepausen mit dir wäre mein PhD nicht halb so spaßig gewesen! Danke für all die tollen Gespräche über leichte und schwierige Dinge, Witz und Verständnis, und unsere gemeinsamen wissenschaftlichen Abenteuer.

Sacha, je hebt zowel professioneel alsook persoonlijk een grote invloed op mijn PhD gehad. Zonder jouw werk was een groot deel van dit proefschrift niet mogelijk geweest. Dank je voor de leukste conferentie reizen die je kan bedenken, ontelbare avonden bij CREA, en dat je mij liet lachen tijdens een van de moeilijkste momenten van mijn leven!

Gaby, dank je voor de leuke conferentie reizen, uitbundige feestjes, en de chille avonden. Ik vind het erg fijn dat we in de laatste anderhalf jaar goede vrienden zijn geworden. Alexander, dank je voor al die tafeltennis potjes en de erg leuke organisatie van het Crossing Borders festival. Jens, danke, dass ich mit dir einen partner in crime in der Sozialpsychologie habe. Florian, danke für all die Gespräche wenn einer von uns beiden mal

wieder an den negativen Seiten der Wissenschaft verzweifelte. Tessa en Riet, dank jullie voor de fijne koffie pauses en leuke gesprekken. Adela, thanks for the great travels and that you always have a cigarette for me. Johnny, dank je voor de interessante gesprekken over muziek en de leuke avonden. Jonas, danke für die spaßigen Treffen und die gemeinsamen rants über measurement error und alles andere was keinen Sinn in der Psychologie macht.

Dear current and past members of the PsychoSystems group, Adela, Angélique, Claudia, Denny, Don, Eiko, Fabian, Gaby, Jolanda, Jonas, Julian, Maarten, Marie, Max, Mijke, Lourens, Pia, Ria, Riet, Sacha, and Tessa, thanks for all the interesting discussions and the great atmosphere. Let's keep on revolutionizing psychology together!

Nicole, danke dafür, dass du als meine große Schwester immer ein Vorbild für mich bist. Bernd, du bist der beste Schwager den man sich wünschen kann. Danke euch beiden für all die schönen Besuche, Spielabende und Skiurlaube. Franca und Miriam, danke für die Leichtigkeit die ihr immer wieder in mein Leben bringt. Es ist eine Freude euch als kleine Schwestern zu haben. Mama, danke, dass du immer an mich glaubst. Papa, danke, dass du immer da bist wenn ich dich brauche.

Andrea, danke, dass du seit nun schon 13 Jahren an meiner Seite bist. Mit dir ist mein Leben sonniger, spannender und leichter. Danke, dass du in mir den Menschen siehst der ich sein möchte!