universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Collective Dynamics of Multi-Agent Networks: Simulation Studies in Probabilistic Reasoning"

verfasst von / submitted by

Max Pellert, Bakk.rer.soc.oec.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (M.Sc.)

Wien, 2017 / Vienna, 2017

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 013
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	MEi:CogSci - Cognitive Science
Betreut von / Supervisor:	Dipl.-Ing. Dr. Paolo Petta

Abstract

The main contribution of this thesis is an empirical evaluation of IBE*, a synthesis of Jeffrey Conditionalization and IBE (Inference to the Best Explanation) that generalizes explanationist updating to cases of uncertain evidence. This thesis furthers our line of research that earlier led to the original development of IBE* in joint project work with Borut Trpin. First, the topic and related core concepts are situated in the broader context of Cognitive Science. The classical paradigm in Cognitive Science (the 'Physical Symbol System Hypothesis') was subject to a 'probabilistic turn'. This is illustrated with a case study. It is argued that although most work done in this new paradigm is Bayesian, there are merits to be expected from studying (probabilistic) alternatives to it. The 'Alien Die' model is introduced and a version with biased coins used subsequently is discussed. The Bayesian formula, a formalization of 'Inference to the Best Explanation' (IBE) and versions that can work with uncertain evidence are described. *Brier scores* are introduced as means to compare prediction accuracy. A flowchart visualizes the typical simulation workflow. Simulations are then run with *full* certainty of evidence. We succeed in replicating a recent key finding by Igor Douven: The explanationist, who adopts IBE as an updating strategy, is faster in assigning high probability to the true bias. In doing so it incurs a slightly higher Brier score. The effect of a specific threshold value on speed and the number of simulations that cross this threshold is also investigated. Next, simulations with *fixed* (un-)certainty for all simulations and all tosses is introduced. The explanationist is again faster and furthermore achieves (substantially) lower Brier scores than the Bayesian alternative. In addition, it is also more *accurate*, i.e. it more often assigns probability above threshold to the true bias. Finally, we analyse simulations with *random* uncertainty. Although higher for both updaters than in the fixed uncertainty case, Brier scores for the explanationist remain the lowest. Likewise, the explanationist remains the significantly faster variant, and is also substantially more accurate: We find a decisive shortcoming of the Bayesian approach. Building on these findings, we can argue that IBE* seems to be counteracting the problem of uncertain evidence. In addition to already named performance advantages, its adoption helps limiting *false positives*, i.e. assignment of probability above threshold to wrong biases.

At this stage, we introduce networks. One by one, random networks, small world networks, BA networks, and q-networks are discussed and visualized under different parametrisation. Matrix representations of networks are used to compute influence matrices. Markov chain theory helps identify under which circumstances influence matrices converge in the limit. For this process, a natural interpretation in opinion dynamics is available. Finally, we characterise the mechanism of *collective* belief updates. We then compare collective belief updates on different network topologies. With full certainty, the explanationist crosses the threshold for the true bias on all topologies. The Bayesian does so only on two out of four topologies. With a distribution of uncertainty of evidence across all agents, this advantage is again more pronounced. We found a vast speed advantage of the explanationist also in the collective scenario for every network topology examined. The last sections contextualise the preceding work: Controversies regarding the choice of specific parameters and concepts used in our simulations are addressed. We develop a definition of computer simulations that allows us to set the stage for a reflection on epistemological issues. In a discussion of the social sciences, we point out the potential of simulations in this domain. Finally, current limitations and further directions of our research are identified.

Zusammenfassung

Der zentrale wissenschaftliche Beitrag dieser Masterarbeit ist die empirische Evaluation von IBE*, einer Synthese aus 'Jeffrey Conditionalization' und des 'Schlusses auf die beste Erklärung' (engl. 'Inference to the Best Explanation', im Weiteren IBE). Dies erlaubt, auch Szenarien unsicherer Evidenz mit 'explanationist belief updates' zu erfassen. Damit schließt diese Masterarbeit an Forschungsarbeiten an, die zuvor zu der Entwicklung von IBE* (gemeinsam mit Borut Trpin) geführt haben. Zuerst, werden die verwendeten Kernkonzepte in den allgemeinen Kontext der Kognitionswissenschaft eingebettet. Das klassische Paradigma der Kognitionswissenschaft (die 'Physical Symbol System Hypothesis') erfuhr eine probabilistische Wende. Ein Fallbeispiel illustriert dies. Der überwiegende Großteil der Arbeiten in diesem Bereich ist bayesianisch. Es wird argumentiert, dass es fruchtbar ist, auch (probabilistische) Alternativen dazu zu untersuchen. Das 'Alien Die' Modell wird in einer Version mit verschieden (un-)fairen Münzen vorgestellt. Die Bayessche Formel, eine Formalisierung von 'IBE' und Versionen davon, die mit unsicherer Evidenz arbeiten können, werden beschrieben. *Brier Punkte* werden verwendet, um die Vorhersagegenauigkeit zu vergleichen. Ein Programmablaufplan zeigt einen typischen Durchlauf der Simulationen. Simulation mit *vollständig sicherer* Evidenz werden dann ausgeführt. Es gelingt, ein zentrales Result von Igor Douven zu replizieren: Der 'explanationist', der IBE als 'updating' Strategie verwendet, ist schneller darin, der wahren Verzerrung der Münze hohe Wahrscheinlichkeit zuzuschreiben. Dabei fährt der Agent aber etwas höhere Brier Punkte ein. Zusätzlich wird der Effekt einer Reihe von verschiedenen Schwellenwerten auf Geschwindigkeit und die Anzahl an Simulationen über diesem Wert untersucht. Im nächsten Schritt werden Simulation mit einer *fixierten Unsicherheit* der Evidenz ausgeführt. Der 'explanationist' ist wieder schneller und erhält dabei deutlich geringere Brier Punkte als die bayesianische Alternative. Zusätzlich ist der 'explanationist' auch *treffsicherer*, d.h. der Agent schreibt hohe Wahrscheinlichkeit (über

dem Schwellenwert) öfter der wahren Verzerrung zu. Schließlich führen wir Simulationen mit *zufälliger* Unsicherheit der Evidenz durch. Brier Punkte sind höher als in den bisherigen Szenarios für beide Typen von Agenten, jedoch immer noch niedriger für den 'explanationist'. Wieder ist dieser Agent signifikant schneller und genauer: Wir finden damit eine entscheidende Schwachstelle des bayesianischen Agenten. Ausgehend von diesen Resultaten können wir argumentieren, dass IBE* dem Problem der unsicheren Evidenz entgegenarbeitet. Zusätzlich zu den bereits genannten Vorteilen, schreibt IBE* weniger oft hohe Wahrscheinlichkeit der falschen Verzerrung zu und liegt damit weniger oft falsch positiv.

Als nächsten Schritt bringen wir Netzwerke in die Analyse ein. Nacheinander, werden 'random networks', 'small world networks', 'BA networks' und 'q-networks' vorgestellt und unter verschiedenen Parametern visualisiert. Repräsentationen dieser Netzwerke in Matrizenform werden verwendet, um 'Einflussmatrizen' zu berechnen. Die Theorie der Markow-Ketten erlaubt es festzustellen, unter welchen Bedingungen diese Matrizen im Limit konvergieren. Der Konvergenzprozess der Einflussmatrizen kann als Meinungsdynamik interpretiert werden. Schließlich charakterisieren wir den Vorgang kollektiver 'belief updates'. Kollektive 'belief updates' werden dann auf den verschiedenen Typen von Netzwerken durchgeführt. Bei vollständiger Evidenz überschreitet der 'explanationist' den Schwellenwert bei jedem Netzwerktypen. Der bayesianische Agent erreicht dies nur bei zwei von vier Netzwerktypen. Wird die Evidenz unsicher, in einer Verteilung über alle Agenten, treten die Vorteile des 'explanationist' stärker hervor. Wir finden einen sehr großen Geschwindigkeitsvorteil des 'explanationist' für alle Netzwerktypen in allen Szenarios mit kollektiven 'belief updates'. Die letzten Kapitel kontextualisieren die vorangegangene Arbeit: Kontroversen bezüglich der Wahl der verwendeten Parameter und Konzepte in den Simulationen werden behandelt. Wir erarbeiten eine Definition von 'Computer Simulationen' und sprechen epistemologische Aspekte an. In einer Diskussion gehen wir auf das Potential ein, das Simulationen für die Sozialwissenschaften besitzen. Schließlich erwähnen wir derzeitige Limitierungen unserer Arbeit und besprechen Möglichkeiten, die Forschung weiterzuführen.

Contents

Acknowledgements

This thesis builds partly on the outcome of the project “Simulation Studies in (Bayesian) Inference”, jointly carried out with Borut Trpin (University of Ljubljana). Thanks for making my stay in Slovenia this productive and for becoming my good friend, Borut!
Some parts of the code used in individual belief updating are inspired by an R script provided to Trpin & Pellert by Igor Douven (French National Centre for Scientific Research), published as supplement to Douven (2016). In particular, we owe Igor Douven (and his collaborators) a very *efficient* way of implementing belief updates in R, that we adapted to our needs in the project.
For his support, I am especially indebted to Paolo Petta.
Above all, I want to express my special gratitude to Susanne Gutsche.
Ne discere cessa!

1 Introduction

The main contribution of this thesis is an empirical evaluation of IBE*, a synthesis of Jeffrey Conditionalization and IBE (Inference to the Best Explanation) that generalizes explanationist updating to cases of uncertain evidence. This thesis furthers our line of research that earlier led to the original development of IBE* in joint project work with Borut Trpin (Pellert & Trpin, 2017).

We begin the presentation of our work by explicating the placement of our research within the field of Cognitive Science.

1.1 Classical Cognitive Science

The classical paradigm of Cognitive Science was established in the early beginnings of the field of Artificial Intelligence. These efforts, which started in the 1950s, culminated in the formulation of the "Physical Symbol Systems Hypothesis" (PSSH) (Newell & Simon, 1976). It is worth considering this significant theoretical proposal not only because of its immense historical importance, but also because of the clarity of argument developed in justifying it.

> **The Physical Symbol System Hypothesis**
> A physical symbol system has the necessary and sufficient means for general intelligent action. (Newell & Simon, 1976, p. 116)

Symbols can be combined to form *expressions* (also called symbol structures). Processes can operate on these expressions. Vis-à-vis the objects that inhabit the world of a symbol system, expressions have two important functions: 1) They *designate* objects: Through such designation, a system can interact with the object. 2) A system can *interpret* expressions: Should an expression designate a process, the system can *carry out* that process (Newell & Simon, 1976, p. 116).

The systems of interest are *physical* to the extent that they obey the laws of physics: Therefore, they should be realizable by engineering efforts. The use of this term also signifies that the proposal goes beyond purely "human" symbol systems: *Necessary* refers to the view that a physical symbol system is *the* prerequisite for *any* agent to show general intelligence. In the analysis, the conclusion is reached that every systems capable of displaying intelligence must be a physical symbol system. *Sufficient* means that one does not need to look any further: A physical symbol system (of sufficient size) can by itself show general intelligence, without the need to add anything to it.

The origin of the hypothesis lies in the efforts of formalizing logic that started towards the end of the 19th century. In this tradition, logic is seen as “a game played with meaningless tokens according to certain purely syntactic rules” (Newell & Simon, 1976, p. 117). From the 1930s onwards, automata theory was developed, among others, by Turing and Von Neumann. This led to the successful construction of physical machines like “ENIAC”, guided by the theoretical blueprint of the novel *science of computers*[1]. Then, suitable programming languages, e.g. LISP (McCarthy, 1978), were created. Finally, the combination of these developments allowed for testing of PSSH: Newell and Simon admit that they know of “no way of demonstrating the connection between symbol systems and intelligence on purely logical grounds” (Newell & Simon, 1976, p. 118). Therefore, the authors’ resort to listing empirical evidence in support of PSSH: On the one hand, attempts at mimicking human performance were successful in specific, restricted domains such as chess. On the other hand, systems were developed that are geared towards generality. The author’s own “General Problem Solver”[2] (GPS) (Newell & Simon, 1963) is one of the first examples of this approach. The design of this physical symbol system was informed by data collected in psychological experiments with probands that were thinking aloud during solving of formal problems. According to the authors, gathering evidence in support of the claim of sufficiency of the PSSH is part of the field of Artificial Intelligence. To justify the claim of necessity, evidence from (cognitive) psychology is needed.

The start of the scientific discipline of AI is conventionally constituted by the Dartmouth Symposium held in the summer of 1955 (McCarthy, Minsky, Rochester, & Shannon, 1955/2006). From our discussion so far it is visible that early AI researchers explicitly introduced at least two of the core disciplines of the cognitive sciences (computer science and psychology). In addition, also linguistics, economics and neurology also played already a role [cf. points 2 & 3 in McCarthy et al. (1955/2006);Miller (2003); and

[1] The term in its modern sense is used the first time in print by Fein (1959). That author argues for the need to establish a new institution, a “Graduate School of Computer Sciences” at American Universities. He points out that the discipline has matured enough to do so and identifies traits relevant to an academic field:

1. The terminology has been established; a glossary of terms exist.
2. Workers in the field do nonroutine intellectual work.
3. The field has sometimes been axiomatized.
4. The field is open, i.e., problems are self-regenerating.
5. There is an established body of literature, textbooks exist, sometimes treatises—even handbooks; also professional journals.
6. University courses, sometimes departments and indeed schools are devoted to the field. (Fein, 1959, p. 11)

Fein also states that “‘[c]omputer science’ is *not* isolated; it is inter-disciplinary” (p.11; italics in original).

[2] The nomenclature also portrays the optimism of the first era: In a classic polemic critique, Drew McDermott already noted in the 1970s: “By now, ‘GPS’ is a colorless term denoting a particularly stupid program to solve puzzles. But it originally meant ‘General Problem Solver’, which caused everybody a lot of needless excitement and distraction. It should have been called LFGNS—‘Local-Feature-Guided Network Searcher’ ” (McDermott, 1976, p. 4).

Simon's background as an economist]. In a way, the birth of cognitive science resulted from a loose disengagement from AI and its practical, engineering approach towards (even) more philosophically inclined basic research. Of course, like any offspring, it inherited traditions, among them its classical paradigm: the PSSH.

To summarize: Classical Cognitive Science is using logic to formalize cognition …

1.2 The Probabilistic Turn in the Cognitive Sciences

… or rather to be more precise: "It would not be unreasonable to describe Classical Cognitive Science as an extended attempt to apply the methods of proof theory to the modelling of thought" (Fodor & Pylyshyn, 1988, pp. 29–30).

This blueprint was challenged and eventually dismissed (by some) as too rigid. One objection refers to the use of classical logic. An important property of classical logic is its *monotonicity*, i.e. "if a conclusion follows from some set of premises, then it still follows when additional premises are added" (Oaksford & Chater, 1991, p. 9). This is clearly violated in everyday reasoning, which is most of the time *defeasible*. Consider the following example:

> While we infer that Tweety flies on the basis of the information that Tweety is a bird and the background knowledge that birds usually fly, we have good reasons to retract this inference when learning that Tweety is a penguin or a kiwi. (Strasser & Antonelli, 2016, beginning of Section 1, "1. Dealing with the dynamics of defeasible reasoning")

To capture this type of reasoning, non-monotonic logics have been proposed. D. Makinson (2003) gradually transitions from classic to para-classical and finally non-monotonic logics, thereby providing an accessible overview. Proposals of non-monotonic logics like "default reasoning" have been shown to be able handle many real-world scenarios that escape classical logics (Pelletier & Elio, 1997).

Around the same time the classical paradigm was objected on the grounds of defeasibility, Bayesian approaches became popular in statistics and were subsequently applied to the psychology of human learning and to cognitive neuroscience. This triggered a "probabilistic turn" in cognitive science and theoretical innovation (Douven & Schupbach, 2015a, p. 1). With this approach, defeasible inferences pose substantially less of a problem. Contrary to classical logics, a probabilistic reasoner can refrain from ascribing the extreme values of 0 (never true) and 1 (always true) to a proposition. Instead, there is an infinite continuum of *degrees of belief* in propositions in the interval $]0, 1[$. If evidence should turn up that is in contradiction with a previously held conviction, it is easy to account for that: The adjustment (in this case lowering) of degree of belief is prescribed by the Bayesian formula:

$$
\begin{aligned}
&P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)} \\
&H = \{\text{Hypothesis (a proposition)}\} \\
&E = \{\text{Evidence (in contradiction to H) }\} \\
&P(H \mid E) = \{\text{Posterior (new) belief in H}\}
\end{aligned} \tag{1}
$$

(We provide a more extensive treatment of Bayes' rule in Section 2.2.)

In case of repeated contradictions (or a really fatal one), belief in H may become infinitesimally small. But also with a higher degree of belief than almost 0, it may not be rationally entertained anymore (for more on this issue, see Section 11.1).

Proponents often also ascribe a *normative* role to Bayesian reasoning: This approach amounts to comparing actual human reasoning to a Bayesian standard (a classical example is Tversky & Kahneman, 1974). Systematic deviations from the posited norm can often be identified. This has lead to the identification of a plethora of potential biases in everyday human reasoning (cf. "List of cognitive biases," 2017). But not only healthy subjects can be studied in this way, as we adress in the next section.

1.3 Case Study: Bayesianism and Schizophrenia

This section provides an account of the explanatory power of Bayesian approaches by presenting a case study in (pathological) human reasoning.

1.3.1 Clinical Considerations

Among other factors, diagnosis of schizophrenia is based on symptoms described by the patient that can be divided in either positive or negative symptoms. Negative symptoms are also called "signs". They involve loss of a normal function. Typically, this is the case with reduced speech output (alogia) or loss of motivation (avolition) (Fletcher & Frith, 2009, p. 48). Positive symptoms describe the presence of an abnormal phenomenon, such as delusions or hallucinations.

The positive symptoms of schizophrenia can be interpreted as the result of a defect in the mechanism that controls and limits the contents of consciousness (C. D. Frith, 1979, p. 225). This implies that the concepts of preconscious and/or subconscious experience play a crucial part in the analysis of schizophrenia. The disorder can be characterized as "excessive self-awareness". Patients are aware of (parts) of the complex informational processing that normally goes on below the level of awareness. This implies that not the processing itself, but the conscious experience of it is abnormal. Symptoms of schizophrenia occur when the selective capacity of consciousness breaks down. Ambiguity arises as there are multiple, inconsistent interpretations of absurd events. In further consequence, there arises substantial difficulty of selecting and carrying through an appropriate course of action. This sketch of the emergence of a very special kind of experience of schizophrenic people gives rise to an important question: Is there any way to study these processes systematically? Speaking historically, there was an insistence on a negative answer to this questions for a long time. However, over the last (few) decades there has been sustained development of a positive answer (for an overview, see Fletcher & Frith, 2009). The core of this work lies in the conjecture, that it is possible to give a systematic account of positive symptoms of schizophrenia with the help of probabilistic theory.

By affirming Bayesianism as a normative reference, it is possible to study delusional inferences as divergences from the norm. Hemsley & Garety (1986) list several of the divergences encountered in clinical investigations on patients. These include deficits at the stage of hypothesis formulation. Especially people suffering from disorganised schizophrenia[3] seem to translate experience directly into belief statements, without considering any related hypothesis and evidence. With all types of schizophrenia, the maintenance of untestable hypotheses is often observed. Often, these involve "end of the world" scenarios directed at a specific time point (e.g., the year 2087). Furthermore, there are also divergences at the "information search" stage: Patients can become absolutely certain of the truth of a delusional belief. They do not see any need to search for disconfirming evidence. This was, e.g. the case with a patient who claimed he "was in direct contact with, and could influence, a famous athlete" (Hemsley & Garety, 1986, p. 54). Finally, there can be problems in assessing the likelihood ratio. This can be illustrated with an example: A patient who formed the hypothesis $H =$"*Police is chasing me.*" is receiving evidence $E =$"*There are police cars on the street.*", while walking to the doctor's appointment. She then assigns very high probability (or even absolute certainty) to $P(E|H)$, which seriously increases her posterior probability $P(H|E)$ and strengthens the paranoid belief of being chased by the police. In Bayesian analysis, the patient forgot to assess $P(E|\neg H)$ correctly (or at all) and to compute $P(\neg H|E)$. In

[3] Diagnostic conventions have changed since the publication of Hemsley & Garety (1986). In the recent version 5 of the "Diagnostic and Statistical Manual of Mental Disorders" (DSM V), schizophrenia subtypes were removed from the definition of the disorder, "due to their limited diagnostic stability, low reliability, and poor validity." Instead, "a dimensional approach ... is included ... to capture the important heterogeneity in symptom type and severity expressed across individuals" (APA, 2013, p. 3).

this way, the patient can be modelled as having forgotten to consider the probability of presence of police cars on the street without her being chased by police.

1.3.2 Evidence

The research program assessing the differences in probabilistic reasoning between healthy and schizophrenic persons is ongoing. Kaplan et al. (2016) is a recent example of research conducted in this area. In their study the authors presented numbers that fluctuated slightly in a random manner around an integer to two groups of probands (schizophrenic and not schizophrenic). After some time, the initially fixed integer changed to another value, and subsequent values fluctuated around this new "anchoring point". Probands had to infer at which point the announced "context change" had occurred. This data was compared to a Bayesian benchmark model. It turned out that schizophrenics slightly overestimated the posterior probabilities of context change. They declared a standard fluctuation more readily to be the context change. Control subjects, on the other hand, more readily interpreted small prediction errors as mere noise.

These results have to be treated cautiously as there were only small effects. What was found is far removed from the authors' idea to provide a marker for schizophrenia. Nonetheless, this research shows that there systematic differences in probabilistic reasoning between people with the disorder and those without it could be identified. Generally, differences in general logical reasoning between deluded patients and controls are small in studies. Kemp, Chua, McKenna, & David (1997) formed two groups of people (deluded and not deluded) that had to answer questions from a standardized test. Subjects could select between responses, that were either logically fallacious or valid. It turned out that both groups often committed logical errors by selecting the fallacious answer. Furthermore, some of the examples provided were classified as neutral, while others were selected because of the emotional content: In the time context of the study[4], an emotional example included the mentioning of AIDS, for example. Including such emotional component reduced logical reasoning performance of the subjects (somewhat more so for the deluded group). Overall, findings of this study imply that impairment of logical thinking is not sufficient to explain delusional thinking. On the contrary, it is especially noteworthy that high IQ in schizophrenic patients is often associated with a special set of symptoms. An anecdotal example would be the story of the highly-gifted mathematician John Nash, who immerged himself in a very elaborate delusion (and could escape from it much later as well). High performance in intelligence tests supports the development of *paranoid* schizophrenia: Higher IQ means more capability of devising

[4] The late 1990s with a high number of awareness campaigns and broad media coverage on this topic.

complex explanatory systems for disintegrated facts (e.g. persecution). Schizophrenic people with low IQ would not be able to make this much "sense" of their abnormal experiences. In these cases, negative symptoms like disorganisation would prevail. Along these lines, some authors go so far to even postulate the existence of a different disorder called "superphrenia". It is similar to schizophrenia but "characterised by high pre-morbid IQ and a clinical profile featuring predominantly positive and affective symptoms, few negative and disorganised symptoms, good insight and good global functioning" (Černis et al., 2015, p. 5).

1.3.3 Assessment

One of the most serious arguments against most of the research covered in this subsections comes from pharmacology: (Almost) all schizophrenic patients in studies take medication. Dosages vary wildly, for example: 295 ± 391 mg/day[5] chlorpromazine equivalent dose of antipsychotics in Kaplan et al. (2016, p. 5). This introduces many possible confounding variables.

Nonetheless, it has been shown that Bayesian reasoning provides a benchmark that can be compared with the reasoning of schizophrenic patients (for more examples see the annotated bibliography in Fletcher & Frith, 2009). As the covered examples demonstrates, valuable insights into psychiatric disturbances can be gained by analysing this contrast.

1.4 Alternatives to Bayesianism?

As we can judge from the discussion above, there are merits in adhering to a Bayesian approach. But, to push the issue once more, is to look beyond Bayesianism: Are there any alternatives that are "probabilistic without being Bayesian"[6] (Douven & Schupbach, 2015a, p. 1)?

A recent tradition in philosophy of science (Douven, 2013, 2016; Douven & Romeijn, 2011; Douven & Schupbach, 2015a, 2015b; Douven & Wenmackers, 2015) considers a probabilistic adaption of "Inference to the Best Explanation" (IBE) as a promising alternative. In its original formulation, IBE refers to the following:

> In making this inference one infers, from the fact that a certain hypothesis would explain the evidence, to the truth of that hypothesis. In general, there will be several hypotheses which might explain the evidence, so one must

[5]Mean ± standard deviation

[6]We are interested in staying probabilistic to take advantage of the ability to revise beliefs without being forced to subscribe to a new framework (as in the case of non-monotonic logics).

> be able to reject all such alternative hypotheses before one is warranted in making the inference (G. H. Harman, 1965, p. 89).

IBE is similar to abduction. Nonetheless, the new name was chosen to avoid some of the ambiguity inherent in the many different concepts of abductive inference. According to Harman, IBE offers the possibility to get rid of controversial (enumerative) inductive inference, which can now be treated as a special case of IBE:

> From the fact that all observed A's are B's we may infer that all A's are B's (or we may infer that at least the next A will probably be a B). [Using IBE we are] making this inference whenever the hypothesis that all A's are B's is (in the light of all the evidence) a better, simpler, more plausible (and so forth) hypothesis than is the hypothesis, say, that someone is biasing the observed sample in order to make us think that all A's are B's. (G. H. Harman, 1965, pp. 90–91)

The advantage is gained transparency: We always have to state the total evidence available when justifying a specific IBE. In analogy to the intermediate results sometimes used in a mathematical proof, Harman calls the stepping stones that lead to an IBE, *lemmas*. In the quotation above, a lemma would be knowledge that the sample is unbiased[7]. There are also disadvantages: With adoption of Harman's IBE we relinquish all attempts to develop a formalized mechanism: In the end, it remains unclear where the inferences come from. They originate from a black hole, that may be called "creativity" or "intuition" for example.

The question of how to decide which of the competing explanations is the *best*[8] one, is deliberately left open in the original formulation of IBE by G. H. Harman (1965). In our simulations, we decided to use the frequencies of specific values in the observations so far to decide on the best explanation. We expand on this in the following sections.

THE INDIVIDUAL DOMAIN

2 Individual Belief Updating

2.1 The Simulated Scenario

This and the following sections give an introduction to the scenario that we investigated empirically in simulations. Additionally, we introduce key terminology (for a general

[7]One could conclude so on the basis of a statistical test.
[8]The question of operationality, so to say.

historical introduction to Bayesianism see for example Joyce, 2011).

Our simulation is populated with agents trying to discover the bias of a coin. This setting was inspired by Van Fraassen's statistical model of the *Alien Die* (1989, pp. 163–170). The name derives from the story behind the model: Van Fraassen considers the usual die, which has the same probability of 1/6 for each value $\{1, 2, 3, 4, 5, 6\}$ to be the "human die". His model, however, focuses on fictional dice, which, according to the story, are to be found on an alien planet. In his discussion of naive inductive inference (extrapolation), Van Fraassen chooses this example to circumvent problems of background knowledge: Normally, when we reason about dice tossing, we use information about human dice. In a formal model, this information (e.g. "its six values are all equally likely") should be accounted for. By supposing alien dice, Van Fraassen aims to develop a "clean" model. He outlines the following scenario: Imagine that alien dice are biased towards showing aces after being tossed. The bias can range between the extremes of all aces and no aces, with any increment of sequence in-between. Without being informed about the actual bias, how can we infer the probability that an ace will show up on the next toss from observing tosses?[9] Van Fraassen uses this setting to show the "fundamental flaw of induction": How to decide about what interpolation "does with small increases in data" (pp. 133-134)? He first considers the simple "straight" rule: The observed ratio of events is said to be the real ratio of events, and therefore must be expected. Quickly, a problem occurs. This rule prescribes, for example, to believe that only aces will occur (with absolute certainty, 1/1) after we have observed just one toss of ace. Clearly, a more subtle rule is needed. Van Fraassen argues for the adoption of the Bayesian formula. In his example, the formula specifies how much to add to the probability that aces will show up from now on, after observing a toss showing ace.

Douven (2013) altered and extended the Alien Die Model. Instead of dice, he considers coins biased to showing heads. The possible biases range from 0 to 1, in steps of 0.1. Douven compares two different belief updating strategies: Standard Bayesianism (see Section 2.2) & IBE (see Section 2.4). To do so, he simulates 1000 rounds of 1000 coin tosses. After each coin toss, a belief update according to either strategy takes place. After each update, two scoring rules (log score and Brier score) are used to measure *epistemic accuracy* (Leitgeb & Pettigrew, 2010). According to published results, IBE updating is typically more accurate at various evaluation points along the sequence of 1000 tosses (e.g. after 100, 250, 500, 750 and 1000 tosses). On the other hand, average accrued penalties for inaccuracy over the whole sequence turn out to be lower for the Bayesian belief update strategy (Douven, 2013, pp. 439–440).

[9]This amounts to guessing the right bias.

	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
Bias:	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Prior:	1/11	1/11	1/11	1/11	1/11	1/11	1/11	1/11	1/11	1/11	1/11

Figure 2.1: Coins that may be used, their bias for heads and their prior probability

2.2 Standard Bayesianism

In our formulation there are the same 11 possible biases towards showing heads $(0, 0.1, \ldots, 1.0)$, as in Douven (2013). Agents set out from a uniform distribution, i.e. considering every bias to be equally likely ($P(Bias0)$[10]$=$ $P(Bias1)$ $=$ ... $=$ $P(Bias10) = \frac{1}{11}$; for an illustration, see Figure 2.1). The vector of these probabilities is the so-called *prior*. After setting their priors, agents are confronted with a coin toss, which results in an outcome of either heads or tails. This is called the *evidence*. The coin can either be a standard, 'fair' coin ($P(Heads) = P(Tails) = 0.5$), or one with any of the other biases (for heads). The extreme cases of $P(Heads) = 0$ ($P(Heads) = 1$) would mean that no heads (or only heads) occur. To allow agents to alter their subjective probabilities conditional on the evidence they receive, we need some more information. The *inverse probability* or *likelihood* denotes the probability that either heads or tails show up, under a certain bias (formally: $likelihood_i = P(heads|bias_i)$ with $0 \leq i \leq 10$, $i \in \mathbb{N}$). If a coin is tossed, by necessity, either heads or tails show up (the coin cannot land on its side). Because of this mutual exclusivity: $1 - likelihood_i = P(tails|bias_i)$. The probability of evidence can be rewritten according to the "Principle of Total Probability" (Joyce, 2011, p. 417): $P(Heads) = \sum_{j=0}^{n-1} P(Heads|Bias_j) * P(Bias_j)$ and $P(Tails) = \sum_{j=0}^{n-1} P(Tails|Bias_j) * P(Bias_j)$; again: $1 - P(Heads) = P(Tails)$ with $n = 11$ (total number of biases).

Now, we have all the parts together that are needed for *Standard Bayesian* belief updating:

$$
\begin{aligned}
&P'(Bias_i) = \frac{P(Bias_i)P(Evidence|Bias_i)}{\sum_{j=0}^{n-1}(Pr(Bias_j)Pr(Evidence|Bias_j))} \\
&i = \{0, \ldots, n-1\} \\
&n = 11
\end{aligned}
\tag{2}
$$

[10] $P(\ldots)$ denotes for the *P*robability of ...

2.3 JC: Jeffrey Conditionalization

In Equation 2, the evidence is fully certain: $P(Evidence) = 1$. Therefore, $P(\neg Evidence) = 0$ and can be forgotten about. This assumption was criticized by various philosophers (for a general overview, see Talbott, 2016, Section 6.2). It has been argued that this view of belief updating is at odds with reality. For example, in scientific, evidence-based reasoning, the full certainty of evidence cannot be taken for granted. After all, scientists give up previously accepted evidence again and again. The philosopher Richard Jeffrey formalized these considerations in a Bayesian approach to reasoning with uncertain evidence (cf. Diaconis & Zabell, 1982). He assumes the probability of evidence to be between 0 and 1 ($0 < P(Evidence) < 1$). Consequently, $P(\neg Evidence) = 1 - P(Evidence)$ has to be taken into account in updating: $P(Bias_i) = P(Bias_i|Evidence) * P(Evidence) + P(Bias_i|\neg Evidence) * P(\neg Evidence)$. Thus, our update formula becomes:

$$P'(Bias_i) = \frac{[P(Bias_i|Evidence)P(Evidence)] * P(Evidence|Bias_i)}{\sum_{j=1}^{n}(Pr(Bias_j) * Pr(Evidence|Bias_j))} + \frac{[P(Bias_i|\neg Evidence)P(\neg Evidence)] * P(\neg Evidence|Bias_i)}{\sum_{j=1}^{n}(Pr(Bias_j) * Pr(\neg Evidence|Bias_j))} \tag{3}$$

$$i = \{0, \ldots, n-1\}$$
$$n = 11$$

Remember that H and T are the only possible outcomes, and that they are mutually exclusive. Therefore, they constitute a partitioning of the sample space. If $Evidence = H$ ($Evidence = T$), then $\neg Evidence = T$ ($\neg Evidence = H$).

An agent that is updates its beliefs according to Equation 3 will hence be referred to as *JC* (or simply *Bayesian*, rather than Standard Bayesian).

2.4 IBE: "Simple" Explanationist Updating

In a number of publications (Douven, 2016; Douven & Romeijn, 2011; Douven & Schupbach, 2015a, 2015b; Douven & Wenmackers, 2015), the philosopher Igor Douven has argued in favor of an extension of Standard Bayesianism. He proposes to add considerations of *explanatory goodness* to the existing framework. This is done by adding a function $f(Bias, Evidence)$ to Equation 2 that, on every update, assigns a bonus to the bias that best explains the evidence so far. Formally:

$$P'(Bias_i) = \frac{P(Bias_i)P(Evidence|Bias_i) + f(Bias_i, Evidence)}{\sum_{j=1}^{n}(Pr(Bias_j)Pr(Evidence|Bias_j) + f(Bias_j, Evidence))}$$
$$i = \{0, \ldots, n-1\} \tag{4}$$
$$n = 11$$

In principle, Equation 4 is indifferent to the concrete measure of explanatory adequacy that will be used. For a discussion of some other measures of explanatory goodness, see Section 11.2.

In the present case, we operationalise explanationist updating in the following way: After computing new probabilities using Equation 2, the bias is selected that explains the whole sequence of heads observed so far with the highest probability. Second, this bias is assigned a *bonus*. Following the example of others (e.g. Douven, 2013, p. 432), we have used a bonus value of 0.1 throughout our present work. Then, the vector consisting of the probabilities of each bias is normalized to 1 again. Finally, each round of updating ends with the newly computed vector being set as the new prior.

It can be said, that the explanationist adopts a *frequentist* measure of explanatory goodness.

2.5 IBE*: Explanationist Updating with Uncertain Evidence

The main contribution of this thesis is an empirical evaluation of IBE*, a synthesis of Jeffrey Conditionalization and IBE that generalizes explanationist updating to cases of uncertain evidence. This thesis furthers our line of research that earlier led to the original development of IBE* in joint project work with Borut Trpin (Pellert & Trpin, 2017).

$$P'(Bias_i) =$$
$$\frac{[P(Bias_i|Evidence)P(Evidence)] * P(Evidence|Bias_i) + f(Bias_i, Evidence)}{\sum_{j=1}^{n}(Pr(Bias_j) * Pr(Evidence|Bias_j) + f(Bias_j, Evidence))} +$$
$$\frac{[P(Bias_i|\neg Evidence)P(\neg Evidence)] * P(\neg Evidence|Bias_i) + f(Bias_i, \neg Evidence)}{\sum_{j=1}^{n}(Pr(Bias_j) * Pr(\neg Evidence|Bias_j) + f(Bias_j, \neg Evidence))}$$
$$i = \{0, \ldots, n-1\} \tag{5}$$
$$n = 11$$

An agent that updates its beliefs according to Equation 5 will hence be referred to as

*IBE** (or simply, *explanationist*).

2.6 Brier Scores

Brier scores are an intuitive way to measure the accuracy of probability predictions. The idea originates from meteorology and was used to assess the precision of probabilistic weather predictions. The following Brier score equation relates the predicted probability of a binary event—for example, $\{rain, \neg rain\}$—to the actual outcome, i.e. we compute the mean squared prediction error:

$$\begin{aligned} &BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2 \\ &f_t \ldots \text{ probability that was forecast} \\ &o_t \ldots \text{ actual outcome of the event at instance t} \\ &\text{(0 if it does not happen and 1 if it does happen)} \\ &N \ldots \text{ number of forecasting instances} \end{aligned} \tag{6}$$

(For Equation 6 see "Brier score," 2017, Section "Definition of the Brier score").

In our simulations, the sample space is also binary: $\{Heads, Tails\}$. Therefore, we can use this simplified version of the Brier score for binary events. (The original formulation of Brier (1950) allows for more than two classes of events happening.) Because we compute Brier Scores after every update, we use $N = 1$ and then track the values over all tosses. With the following formula, an adaption of Equation 6, we compute Brier scores after every belief update:

$$\begin{aligned} &\text{BRIER} = (1 - P(Bias_c))^2 + \sum_i (0 - P(Bias_i))^2) \\ &c = \{\text{Index of the true bias}\} \\ &i = \{\text{Indices of the false biases}\} \end{aligned} \tag{7}$$

3 Workflow of Running a Simulation with Typical Parameterization

Figure 3.1 shows the workflow in a simulation with typical parameters.

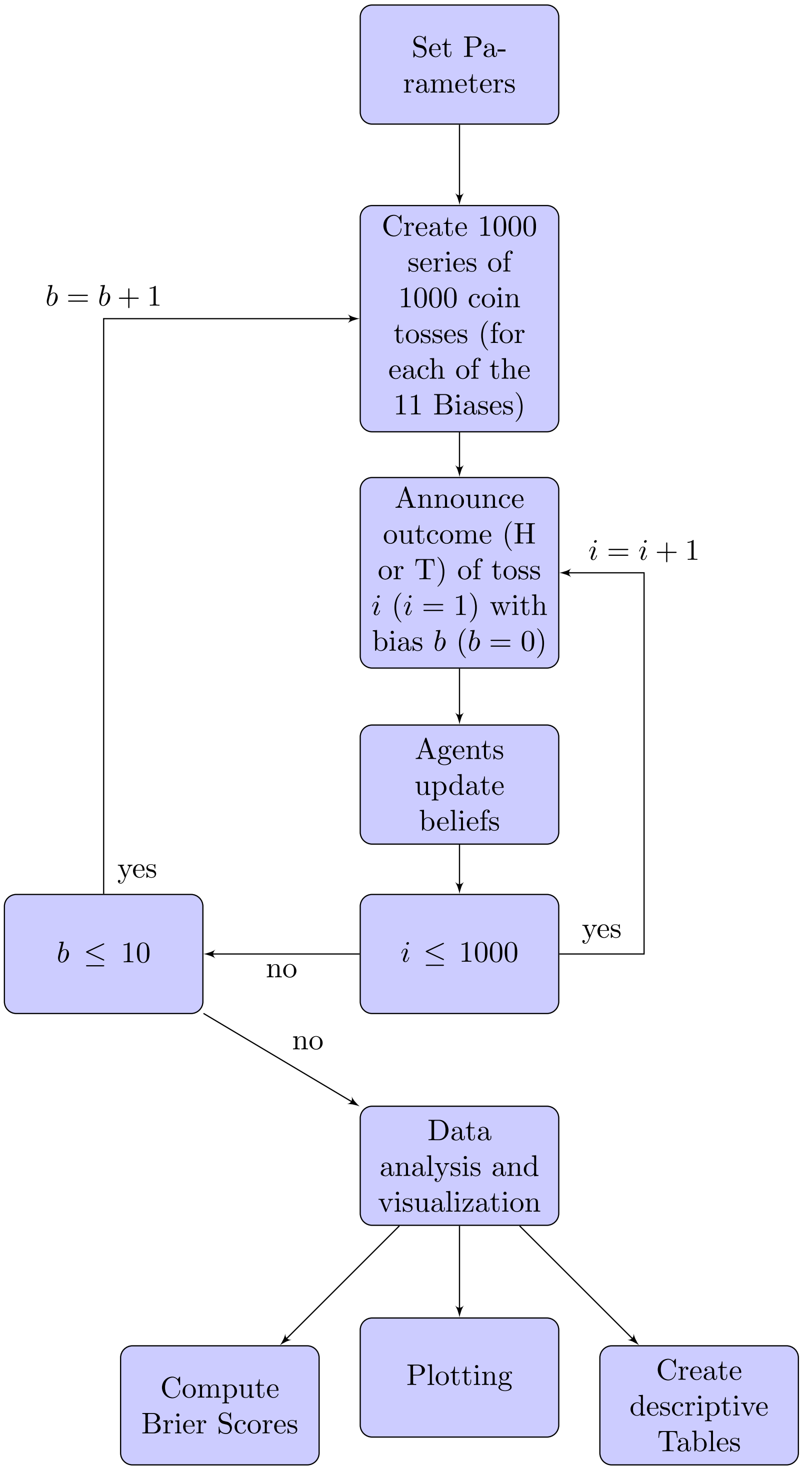


Figure 3.1: Flowchart of the simulations

4 Technical Details

The original simulation was developed in Python. By using the library "PythonInR" in RStudio, it was possible to take advantage of the power of R in statistical evaluation and visualization. The R package "ggplot2" was used extensively to create plots. In a later stage, we were provided with an R source file, used in Douven (2016): By adapting and extending this script, it was possible to shift the simulation part of the project from Python to R. Directly working in R increased efficiency, because the package that allowed to source Python scripts in R proved to be a performance bottleneck. Furthermore, some parts of the new simulation code were rewritten in C++ and integrated into R with the package *Rcpp*. This allowed to shorten time needed for computation even further, by a factor of approximately 0.5: Loops can be pretty costly in R, while in C++ they are not. Especially the bonus assignment function for the explanationist benefited from this change. The R package "parallel" was used in the thesis for some simulation runs. It allows to take full advantage of multicore CPUs. Standardly, R uses only one core even if more (in the author's case, four) cores are available. We decided to use three out of the four cores cores to have one core available for other tasks. In this way, we were able to speed up operations substantially, by a factor of approximately 2.5.

This master thesis was written in RMarkdown, which allows to create documents with Markdown formatting in combination with dynamically created R and LaTeX content.

In the implementation of the mechanism of belief updating in our simulations we take advantage of the"Reduce" function in R (cf. the online documentation "Common Higher Order Functions in R package 'base' version 3.5," 2017). It is modeled after functions commonly found in other programming languages (e.g. "accumulate" in C++: "Accumulate - C++ Reference," 2017).

"Reduce" from the R package "base", takes two vectors and a function as arguments. The first vector consists of a sequence of integers from 1 to *number of coin tosses*. The second vector is the so-called "initialization vector". In our case this is the vector of our uniform priors. The function that is provided must be binary, i.e. it must take exactly two arguments. "Reduce" first feeds it the initialization vector as first argument and the first value of the coin tosses. The vector returned by the binary function is recursively fed to the function itself, as first argument, with the second value of the first argument vector to the Reduce function as second argument. The return vector is again fed as the new first argument, together with the third value of the first argument of the Reduce function. And so on, for all values of the first argument of the Reduce function. The sequence of all vectors returned by the binary functions is returned by the Reduce function (rather than the result of the final invocation), if specified so by the optional additional Boolean “accumulate” argument.

What happens in the call to the binary function? The first argument, the vector, is used as the prior in the belief update procedure. The second argument, an integer, is used as an index to look up the outcome of the actual coin toss, which is the relevant evidence on which belief updating is then performed. Additional tasks are also performed: In the case of the explanationist, a function is called that assigns the bonus. In this way, the Reduce function can be used to implement belief updating according to different strategies.

4.1 Work environment

```
R version 3.4.1 (2017-06-30)
```

```
ggplot2: 2.2.1
igraph: 1.1.2
network: 1.13.0
pander: 0.6.1
Rcpp: 0.12.12
rmarkdown: 1.6
xtable: 1.8.2
```

```
Windows 8.1 x64 (build 9600)
```

```
x86_64-w64-mingw32/x64 (64-bit)
```

5 Simulations With Fully Certain Evidence

In this section we explore a scenario with fully certain evidence, where IBE* updating reduces to IBE and Jeffrey Conditionalization to Standard Bayesianism. Put more formally: $P(Evidence) = 1$ and $P(\neg Evidence) = 0$, which cancels out the second part of Equation 3 & 5.

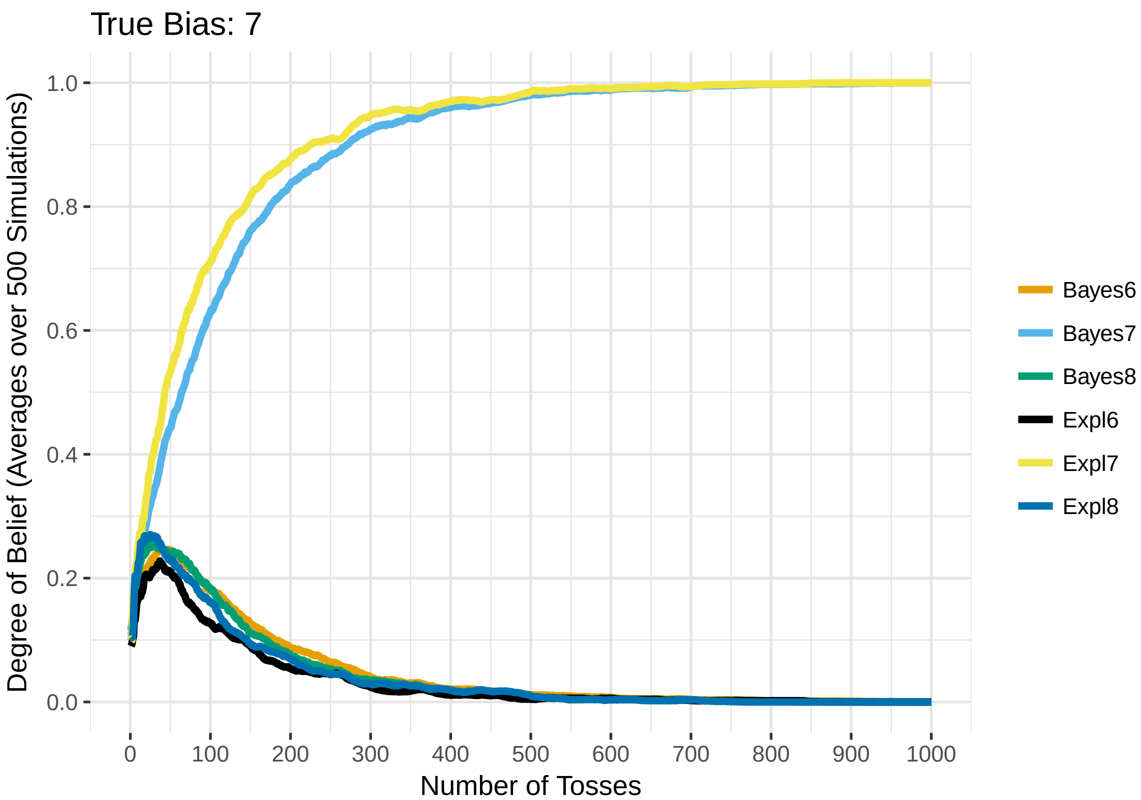
5.1 Parameter Settings

The following settings are used:

- Certainty is fixed at 1 (there is full certainty of evidence).
- Number of simulations: 500.

- To prevent memory problems we decided not to set a higher number (see Section 13.2).
- Number of coin tosses per simulation: 1000.
- Threshold is set at 0.9.
  - Different values are discussed in Section 11.1.
- Bonus is set at 0.1.
  - We stick to this value in accordance with recent literature on IBE (cf. Section 2.4).

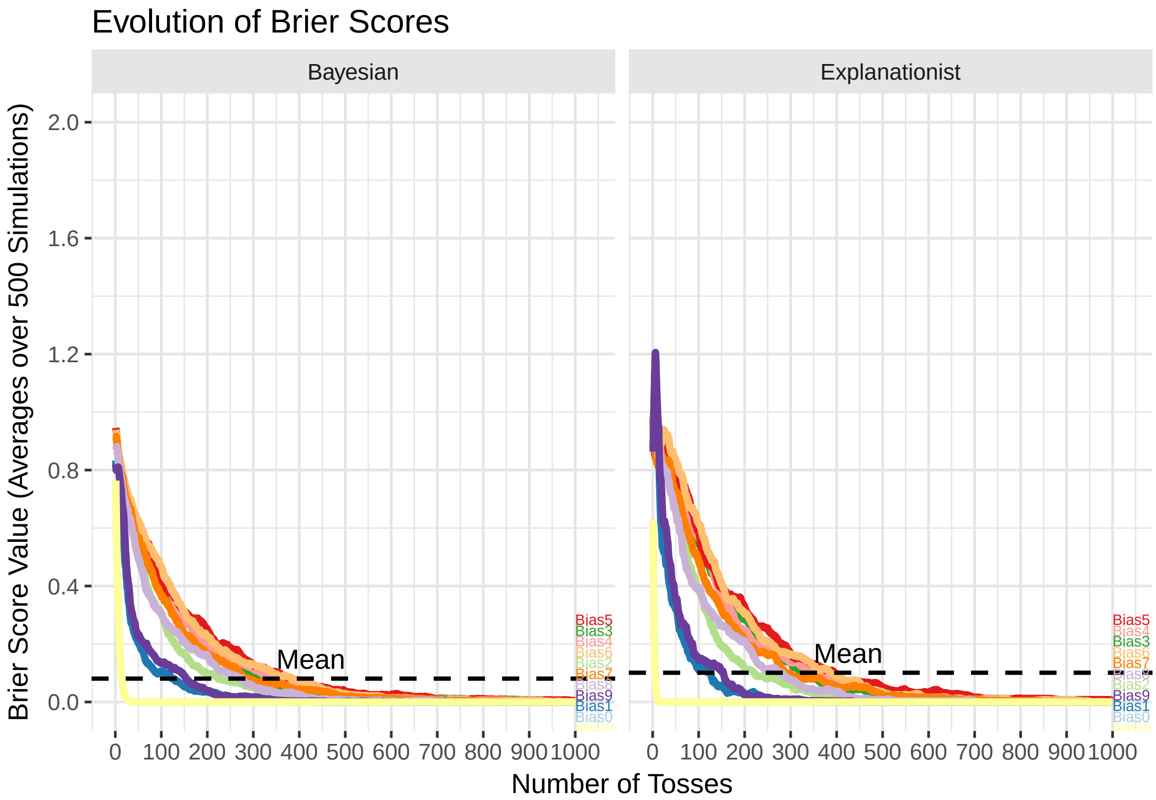
5.2 Aggregate Belief Evolution



This plot shows the replication of an earlier finding (Douven, 2013, p. 433): The explanationist is faster in assigning high subjective probability to the true bias. The explanationist's degree of belief in Bias 7 (yellow line) is steeper for (roughly) the first 400 tosses compared to Bayesian's belief in that bias (light blue line). That means that adopting explanationist updating provides a speed advantage in the case of fully certain evidence compared to Bayesian updating: It takes the explanationist a *smaller number of observed tosses* to form a degree of belief in the true bias above threshold of 0.9, i.e. it faster.

Speed issues are investigated in-depth in Section 5.5 & 6.5.

5.3 Brier Plots



In terms of Brier Scores (i.e. overall accuracy of beliefs), both updaters perform almost equally well. Because the Explanationist is more easily "led in the wrong direction" during the first few tosses, its overall mean is slightly higher (0.1 to 0.08). This divergence in the beginning is due to the sparse information that is available for the bonus function. Naturally, the sequence of coin tosses (the frequencies of heads and tails) has not much explanatory power if it consists only of a few entries. Therefore, wrong hypotheses are easily rewarded after observing only a few tosses, which explains the early (small) spike in Brier scores for the explanationist. This was already noted and discussed by Douven (2013): "[The higher average Brier scores are] due to the fact, noted earlier, that the explanationist reacts in a bolder fashion to the evidence than does the Bayesian and is therefore more easily led astray by rows of tosses which produce subsequences with deviating relative frequencies" (p. 441).

5.4 Performance Tables

As can be seen from the table below, both updaters perform well in assigning a probability equal or higher than 0.9 to any of the biases over the course of 1000 coin tosses:

Table 5.1: Number of Cases With Belief Above Threshold of 0.9 (Fully Certain)

	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
Expl.	500	500	500	500	500	499	500	500	500	500	500
Bayes.	500	500	499	498	499	498	497	500	500	500	500

It turns out that they in almost every case assign the probability to the correct bias:

Table 5.2: Number of Cases With **True** Belief Above Threshold of 0.9 (Fully Certain)

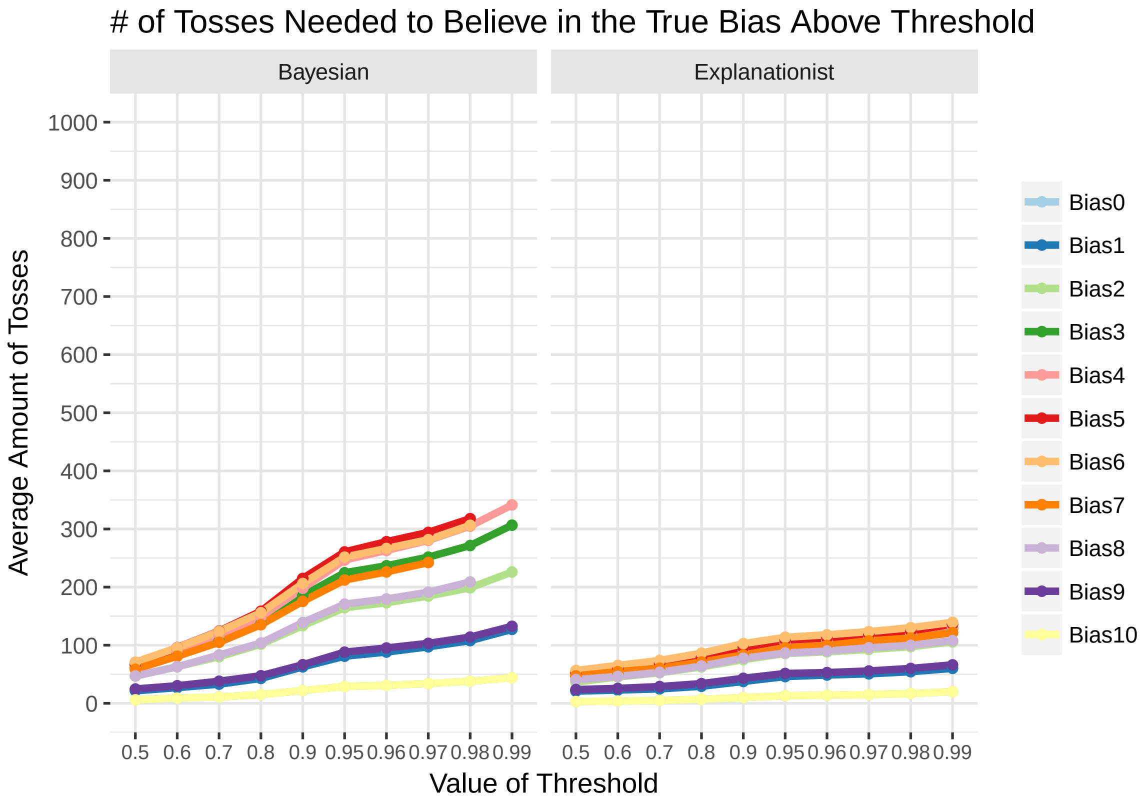
	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
Expl.	500	500	500	499	499	498	500	500	500	500	500
Bayes.	500	500	499	497	499	497	497	500	500	500	500

But, the Explanationist is faster in doing so. It assigns a probability ≥ 0.9 to the true bias on average with 64.7 (SD = 41.2) fewer tosses than the Bayesian. This means that it is on average 1.98 (SD = 0.25) times faster:

Table 5.3: Average Number of Tosses Needed to Form a Belief in the True Bias Above Threshold of 0.9 (Fully Certain)

	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
Expl.	10	37.7	75.2	87.5	90.9	96.0	102.8	83.3	77.6	43.1	10
Bayes.	22	62.8	133.7	185.0	198.0	214.9	206.2	175.5	139.0	66.4	22

5.5 Speed and Threshold (Fully Certain)



Note that we use a non-linear abscissa in order to bring the high thresholds (≥ 0.95) in the focus, while still being able to show the two plots next to each other.

With these plots we show that the Explanationist is faster in detecting the true bias (for every bias) at every threshold. The Bayesian is starting not to manage to cross threshold for the true bias for some of the higher thresholds. This can be inferred from its missing points, e.g. for Bias 7 (light red) at threshold 0.98.

5.5.1 Threshold and Evolution of Belief in Any Bias (Fully Certain)

Table 5.4: Explanationist: # of Simulations Above Threshold (Fully Certain)

Threshold	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
0.50	500	500	500	500	500	500	500	500	500	500	500
0.60	500	500	500	500	500	499	500	500	500	500	500
0.70	500	500	500	500	500	499	500	500	500	500	500
0.80	500	500	500	500	500	499	500	500	500	500	500
0.90	500	500	500	500	500	499	500	500	500	500	500
0.95	500	500	500	500	500	499	500	500	500	500	500

Threshold	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
0.96	500	500	500	500	500	499	500	500	500	500	500
0.97	500	500	500	500	500	499	500	500	500	500	500
0.98	500	500	500	500	500	499	500	500	500	500	500
0.99	500	500	500	500	500	499	499	500	500	500	500

Table 5.5: Bayesian: # of Simulations Above Threshold (Fully Certain)

Threshold	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
0.50	500	500	500	500	500	500	500	500	500	500	500
0.60	500	500	500	500	500	500	500	500	500	500	500
0.70	500	500	500	500	500	500	499	500	500	500	500
0.80	500	500	499	499	500	499	498	500	500	500	500
0.90	500	500	499	498	499	498	497	500	500	500	500
0.95	500	500	499	498	498	498	495	500	500	500	500
0.96	500	500	499	497	498	498	495	499	500	500	500
0.97	500	500	499	497	498	497	495	499	500	500	500
0.98	500	500	499	496	497	496	495	499	500	500	500
0.99	500	500	499	495	497	495	493	498	499	500	500

Both updaters easily get over threshold, even at very high levels. They assign probability higher than the respective threshold in (almost) all cases to any of the biases (not necessarily the true one).

6 Simulations With Fixed Uncertainty of Evidence

In this section we explore a scenario with fixed uncertainty of evidence. The explanationist updates according to IBE* and the Bayesian according to Jeffrey Conditionalization.

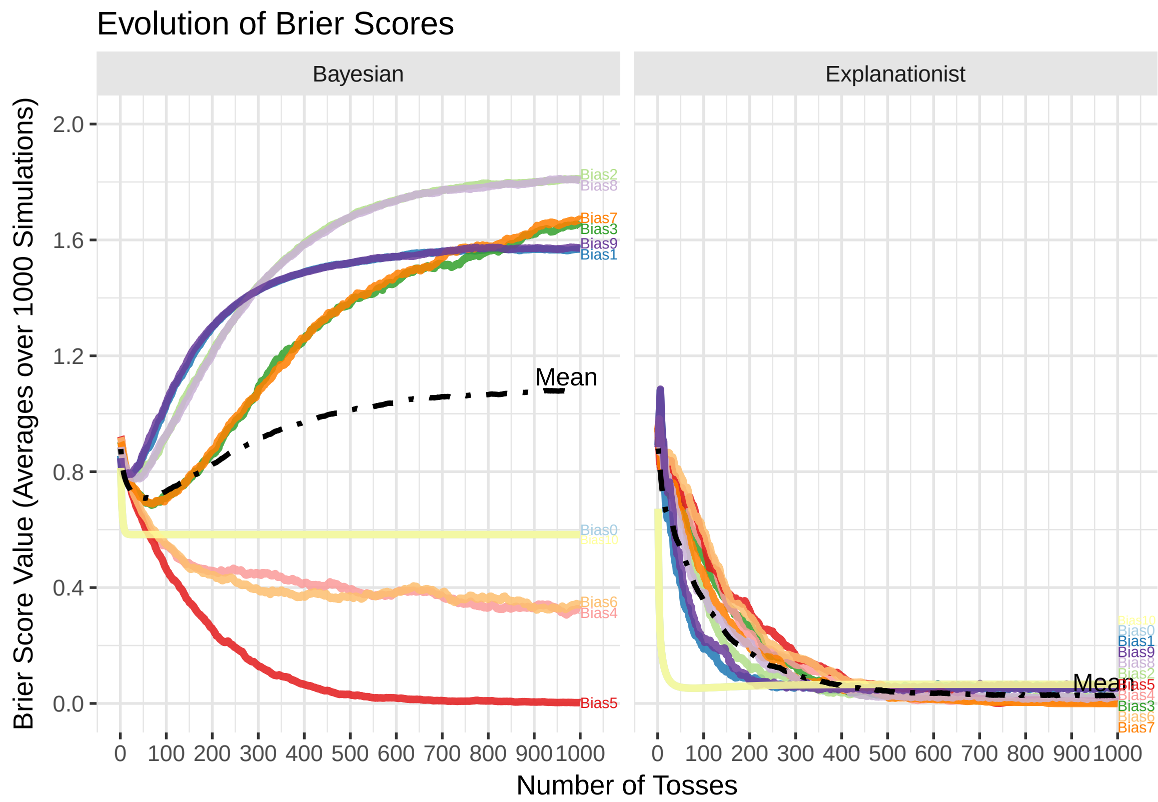
6.1 Parameter Settings

The following parameter is changed, the others remain as before:

- Certainty is fixed at 0.8.

- We move away from full certainty ($P(E) = 1$), but decide not to choose a value below $P(E) = 0.5$: In the case that evidence is less than 0.5 certain, ¬Evidence becomes relatively more certain. If for example, the outcome was heads with 0.3 certainty, it would appear more likely the outcome to have been tails. Such probability of evidence is however misleading and will therefore be excluded from our simulations. We therefore had to decide on a single value in the range of $]0.5, 1[$ for certainty that stays fixed for both agents over all tosses in all simulations. We set $P(E) = 0.8$ which seems far enough removed from both the minimum and the maximum values in the admissible range to provdide a balanced example. Nonetheless, this decision remains arbitrary to some extent.

6.2 Brier Plots

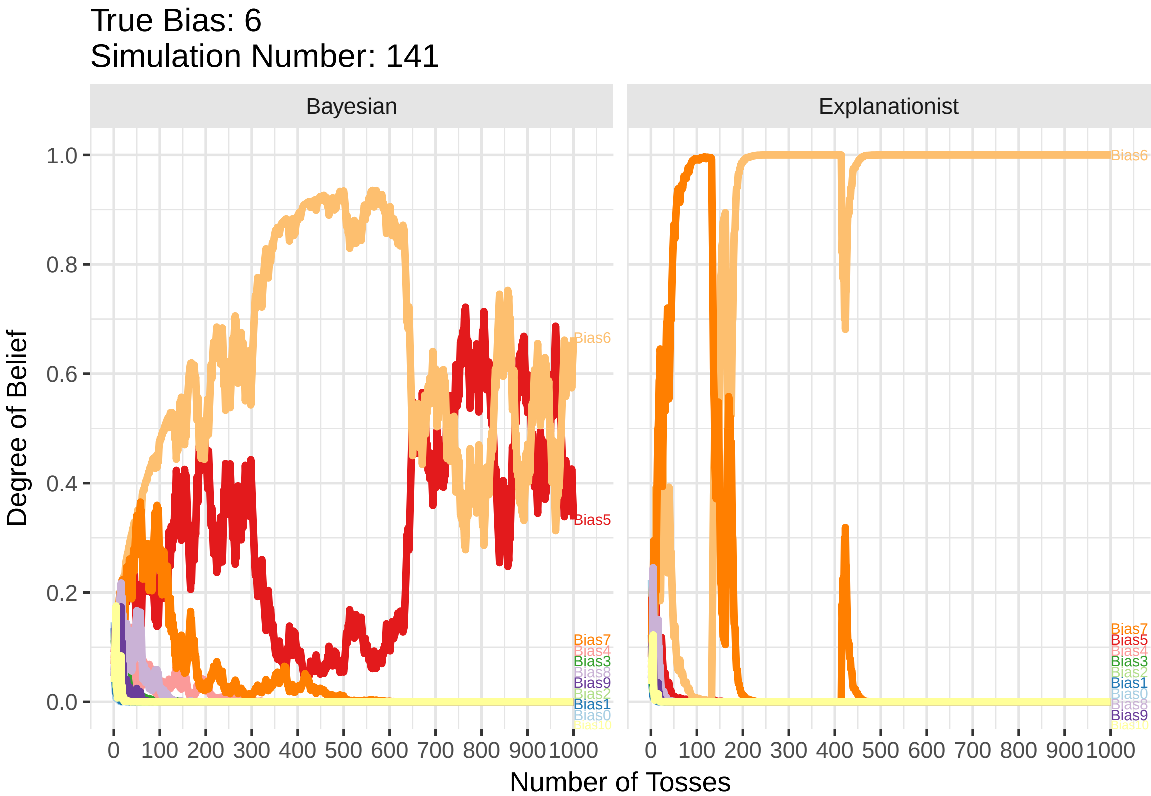


After a slight fall in the beginning, Brier scores for the Bayesian averaged over all biases increase for all subsequent tosses. For the explanationist, they are declining. The overall mean (over all hypotheses, over every toss) of Brier scores is substantially lower for the explanationist (0.12 to 0.96; SD = 0.03 to 0.53). This is a novel finding: Belief updating according to IBE* seems to counteract the problem of uncertain evidence. In contrast to the simulations with full certainty of Section 5, the explanationist is now more accurate than the Bayesian.

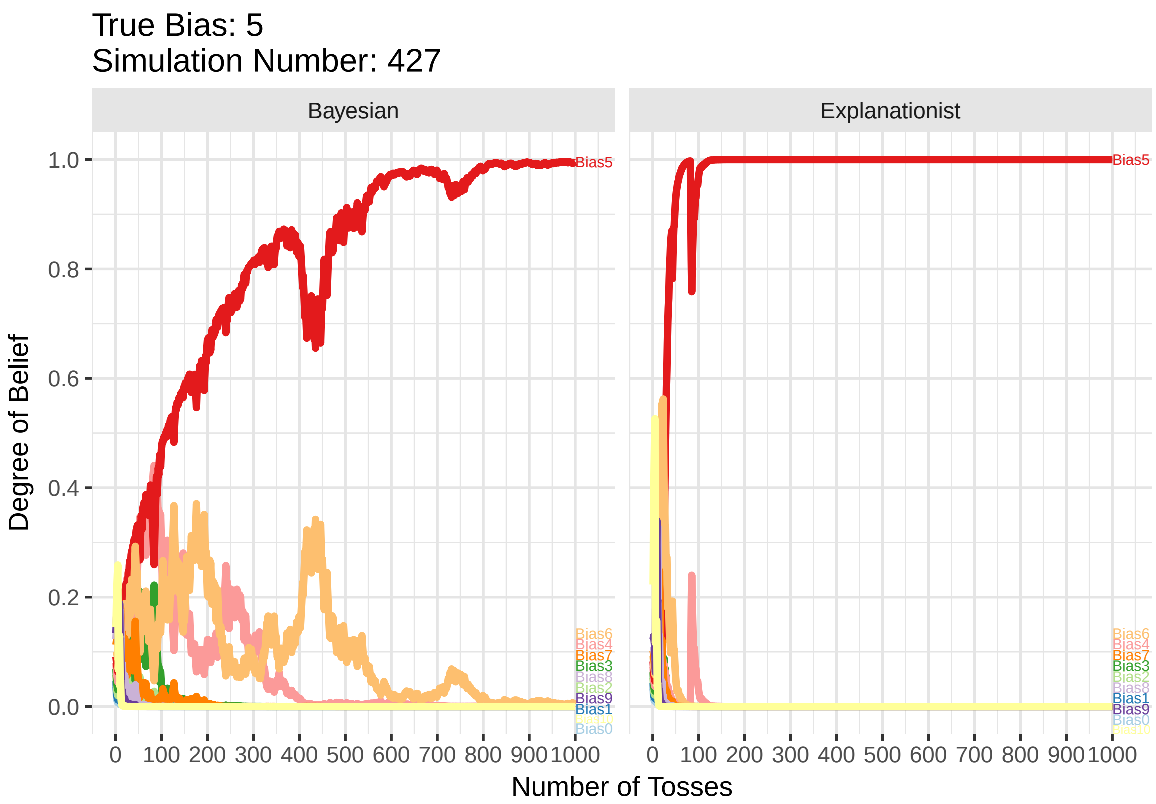
Note the symmetry: In Bayesian updating, biases that lie exactly opposite in the range of possible biases (e.g. B_1 and B_9, B_3 and B_7) incur the same Brier scores.

6.3 Evolution of Belief in Randomly Selected Simulations

In this section, we select random runs out of the 500 simulations. We plot the results to highlight the dynamics of belief updating over 1000 coin tosses. This allows to discuss results that may be lost through averaging over all simulation runs.



The explanationist's lines are steep, i.e. it reacts to evidence in a bold fashion. It crosses threshold for the first time for the *wrong bias*. This can happen precisely because of the boldness of IBE*, when evidence is *not abundant*. Nonetheless, IBE* can correct its mistake when more information becomes available through the subsequent coin tosses and the true bias takes over. The Bayesian is slower. It also manages to cross the threshold of 0.9 for the true bias 6, but the it does so slower and not *persistent*: It starts to assign high probability to the wrong bias 5 and eventually oscillates between these two biases below threshold. The explanationist, on the other hand, sticks with believing in the true bias above threshold.



The explanationist crosses the threshold for the first time with toss 49. It then shortly falls below threshold again. This can be attributed to the still low number of observed tosses at the time, which results in sparse information to be used in bonus assignment. In line with this reasoning, it shortly afterwards crosses threshold persistently with toss 64. The Bayesian does so with toss 491 (and stays there). The explanationist has a vast speed advantage in this simulation run.

6.4 Comparison Tables

The explanationist is believing in any bias above threshold more often than the Bayesian:

Table 6.1: Number of Cases With Belief Above Threshold of 0.9 (Fixed Uncertain)

	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
Expl.	0	3	367	500	500	499	500	500	365	0	0
Bayes.	0	4	274	331	63	492	63	342	271	14	0

The explanationist is much more accurate. By comparing Table 6.1 and 6.2, one can see that the Bayesian often runs the risk of assigning high probability to the wrong bias. The explanationist is much less prone do so. In fact, if it assigns probability above threshold, in the vast majority of simulations, it does so to the true bias:

Table 6.2: Number of Cases With **True** Belief Above Threshold of 0.9 (Fixed Uncertain)

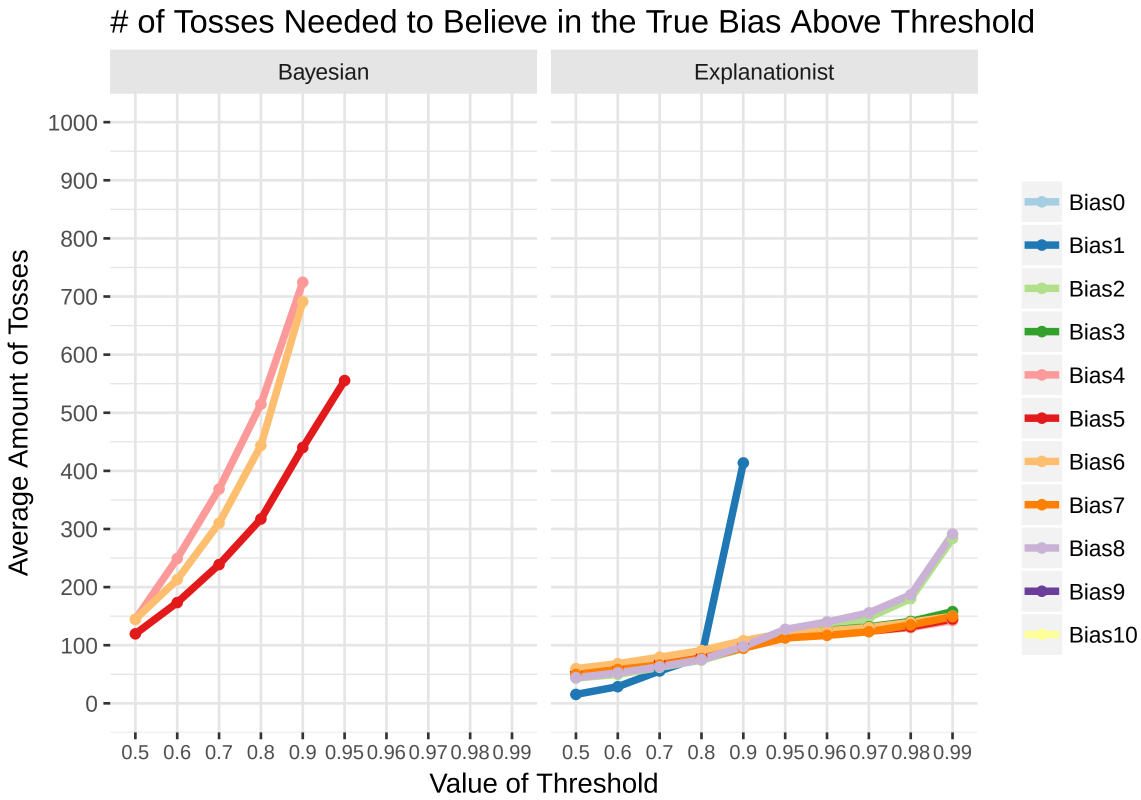
	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
Expl.	0	3	367	499	499	498	500	500	365	0	0
Bayes.	0	0	0	0	62	492	58	0	0	0	0

Not only is the explanationist much better at detecting the right bias at all, it is also on average much *faster* in doing so. For the few biases (B_4, B_5, B_6), where the Bayesian reaches the threshold for the true bias, the explanationist assigns a probability of 0.9 on average with 516.1 (SD = 154.5) fewer steps than the Bayesian. This means it is on average 6.04 (SD = 1.52) times faster:

Table 6.3: Average Number of Tosses Needed to Form a Belief in the True Bias Above Threshold of 0.9 (Fixed Uncertain)

	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
Expl.	0	413.7	94.5	101.6	98.8	100.7	107.8	94.8	97.9	0	0
Bayes.	0	0.0	0.0	0.0	724.5	440.1	691.1	0.0	0.0	0	0

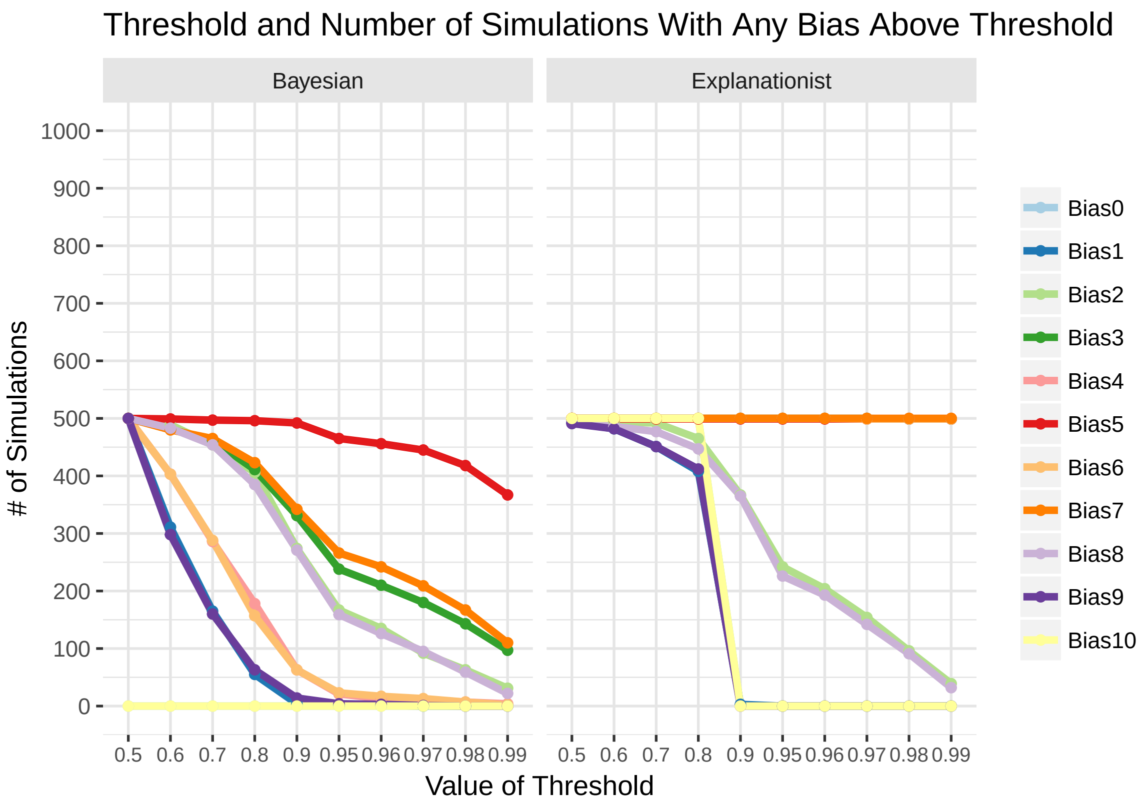
6.5 Speed and Threshold (Fixed Certain)



Note that we use the same non-linear abscissa as in Section 5.5.

The steepness of the Bayesian's lines show its severe weaknesses: It is slow compared to the explanationist. Additionally, missing dots and lines indicate again that the Bayesian fails to cross the threshold for the true bias (even for quite low values).

6.5.1 Threshold and Evolution of Belief in Any Bias (Fixed Certain)



The plot above shows that the number of simulations with any bias above threshold decreases quicker with increasing threshold value for the Bayesian. The explanationist passes the threshold for a high number of simulation until a threshold value of 0.8 and we see decreasing numbers of simulations afterwards. For the Bayesian there is steady decrease starting already with a threshold of 0.5.

6.6 Evaluation

In this section, uncertain evidence was introduced. Certainty was fixed at 0.8 for both updaters in all tosses and all simulations. We found that the explanationist can increase its lead over the Bayesian in terms of various dimensions of performance. It manages to cross threshold more often and does so, most of the time, for the true bias. It is faster and contrary to the simulations with full certainty of evidence, it does incur lower Brier penalties.

By adopting an explanationist updating strategy it also seems possible to limit *false positives*: The Bayesian often believes in a wrong bias above threshold. If the explanationist, on the other hand, believes in a bias above threshold, it is much more often the true one (see Table 6.1 and 6.2). The relevance of this feature depends on the context: In medical diagnosing, it is very valuable, given the grave consequences of errors in this domain. In a scenario in which any decision for action is better than being passive (principle of “action is required”) more importance may be placed on reaching the threshold for any bias as often as possible. Generally, there is a trade-off between being *bold* (more often threshold reached, less often for the true bias) and being *correct* (less often threshold reached, more often for the true bias). The validity of this finding depends on the domain: We are investigating a scenario, where evidence is *abundant*. In scenarios with scarce evidence, e.g. observation of rare events, IBE* may not help in limiting false positives.

In general, measures of explanatory goodness in belief updating strategies seem to help to counteract the problem of uncertainty of evidence. To investigate whether this finding holds for values of certainty other than 0.8(as investigated in this section), we investigate a scenario with *random uncertainty* in the next section.

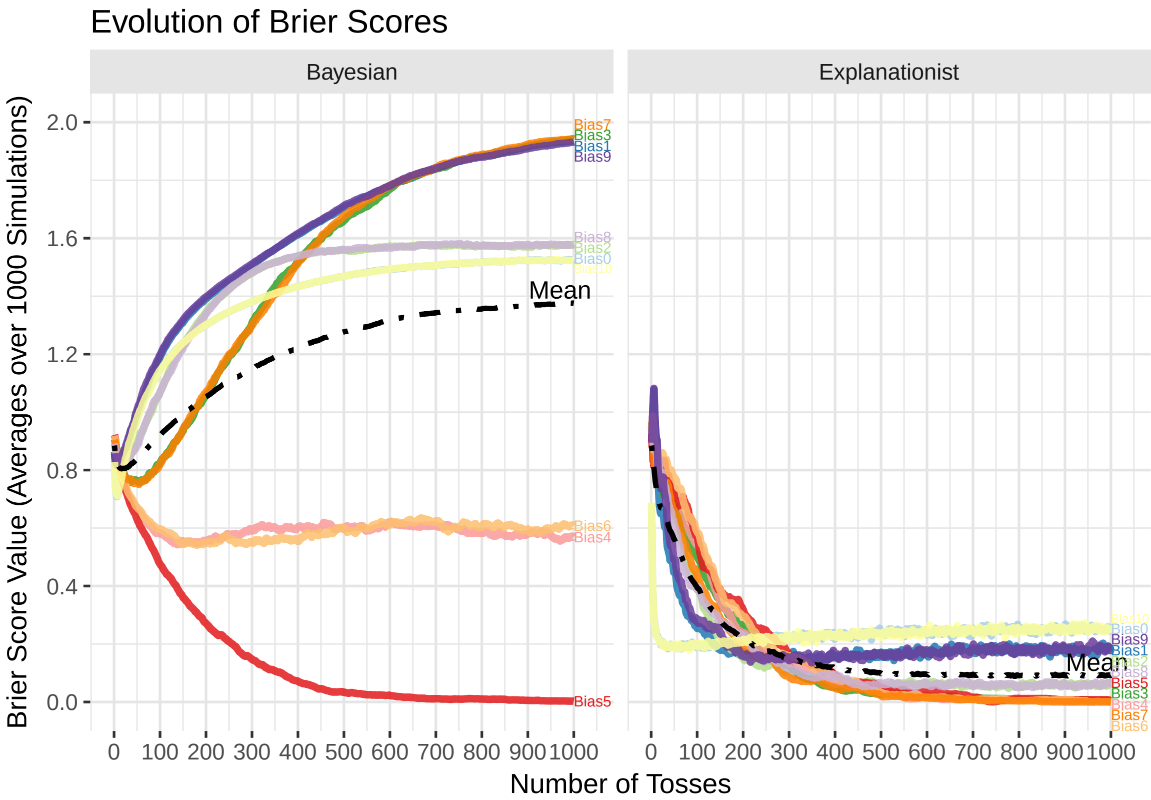
7 Simulations with Random Uncertainty of Evidence (RT+)

7.1 Parameter Settings

The following parameter is changed, the others remain as in Section 5:

- Uncertainty of evidence is established **R**andomly on every coin **T**oss in the range of $]0.5, 1[$. This means that evidence changes randomly, but it is not misleading (cf. Section 6.1).

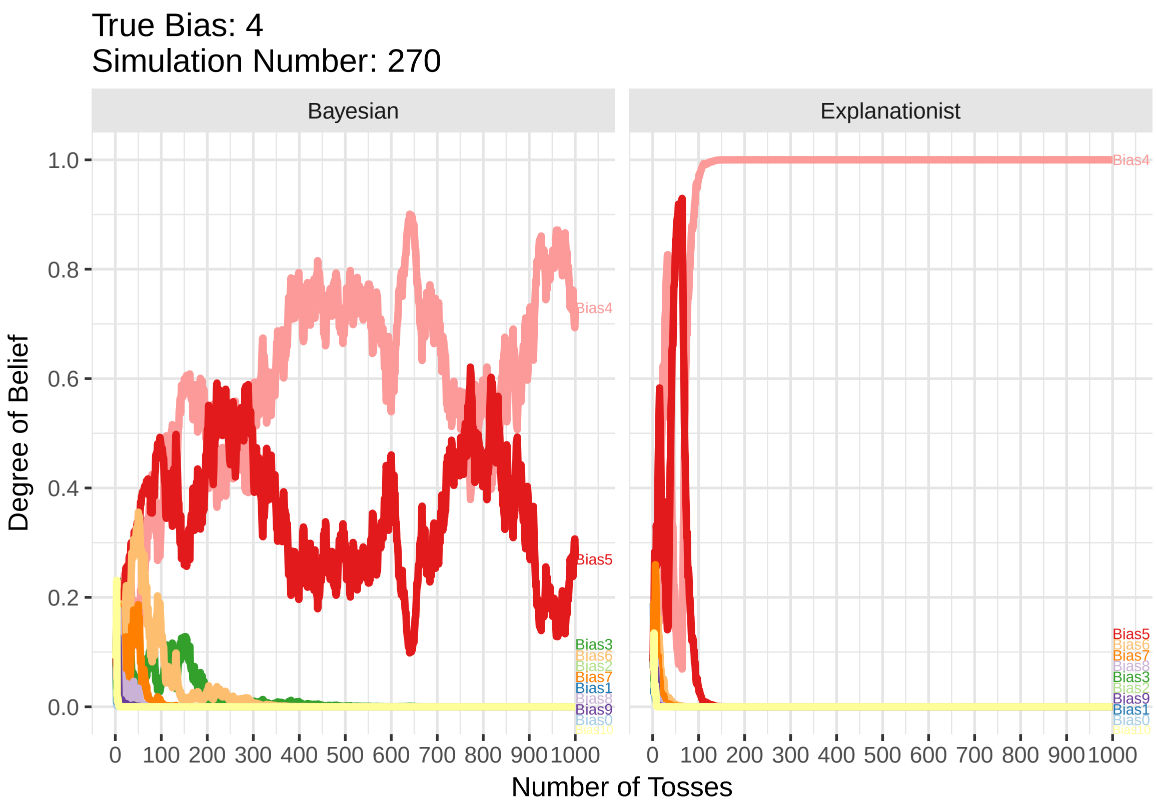
7.2 Brier Plots



For the Bayesian, different biases lead to a different Brier score evolution. In detecting some biases it does not accumulate high Brier values, in others (the majority) it does. For the explanationist, Brier scores decrease for all biases and settle around slightly different (low) values. Compared to the case of fixed uncertainty of evidence, Brier scores are higher for both updaters. Overall, the explanationist still has a much lower average Brier score (0.17 to 1.21; SD = 0.04 to 0.51).

7.3 Evolution of Belief in a Randomly Selected Simulation (RT+)

As in Section 6.3, we again selected random runs out of the 500 simulations. This allows to examine the effect that the introduction of random uncertainty has on belief updating in a single simulation.



Random uncertainty introduces more volatility in the beginning of the simulations. After initially believing above threshold in the wrong B_5, the explanationist concludes on the right B_4. The Bayesian just slightly crosses threshold of 0.9 for the true bias at toss 640. After that, there is a match between B_4 and B_5 for a few hundred tosses. After 1000 coin tosses, it settles on the true B_4, but under threshold.

7.4 Comparison Tables

The full and almost full biases (B_0, B_1, B_9, B_{10}) do pose problems, but overall the explanationist is quite often crossing threshold for any bias. The Bayesian does so in a smaller number of simulations:

Table 7.1: Number of Cases With Belief Above Threshold of 0.9 (Random Certain)

	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
Expl.	6	42	217	500	500	499	500	498	226	37	6
Bayes.	0	476	11	483	28	485	27	480	13	472	0

Here, the Bayesian fails decisively. Except for B_5 (and for some cases of B_4 and B_6) it always believed in the false bias, when it managed to believe in any bias at all. This is in sharp contrast to the explanationist: In all but 3 simulations, it believed in the true bias, in case it managed to cross threshold:

Table 7.2: Number of Cases With **True** Belief Above Threshold of 0.9 (Random Certain)

	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
Expl.	6	42	217	499	499	498	500	498	226	37	6
Bayes.	0	0	0	0	16	485	10	0	0	0	0

It takes the explanationist a little bit longer than in the scenario with fixed certainty to reach the threshold (averaged over all biases: 100.9 to 128.5 tosses). The Bayesian is very slow in forming a belief in the true bias above threshold for the few biases (B_4, B_5, B_6), where it actually manages to do so. With random uncertainty, the explanationist assigns a probability of 0.9 on average with 582.2 (SD = 181.7) fewer tosses than the Bayesian for B_4, B_5 & B_6. This makes the Explanationist on average 6.61 (SD = 1.65) times faster:

Table 7.3: Average Number of Tosses Needed to Form a Belief in the True Bias Above Threshold of 0.9 (Random Certain)

	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
Expl.	263	123.9	103.5	105	101.1	100.7	108.4	101.9	108.9	114	183
Bayes.	0	0.0	0.0	0	777.0	473.4	806.3	0.0	0.0	0	0

7.5 Summarizing Findings

The simulations of Sections 5, 6 and 7 show that it can be rewarding to go beyond a Standard Bayesian belief updating strategy. IBE, an alternative strategy in the case of fully certain evidence, allows for faster belief formation. If evidence is uncertain, IBE* is superior to (Standard) Jeffrey Conditionalization in terms of speed *and* accuracy. With randomized uncertainty of evidence, these advantages are even more pronounced.

THE COLLECTIVE DOMAIN

8 Introducing Networks

The systematic study of networks (conventionally called “graphs”) has a long history in mathematics. Leonhard Euler’s analysis of the problem of the “Seven Bridges of Königsberg” in 1736 is seen as the starting point of graph theory. Twenty years ago, modern “network science” began to boom. It has caused a lot of euphoria since and is sometimes proclaimed to be the “next scientific revolution” (Barabási, 2002, p. 8).

In the following, different types of networks are presented. The number of vertices and edges is kept at a certain limit (typically 20), because static visualizations of bigger networks tend to become unintelligible hairballs.

8.1 Random Networks

Erdős & Rényi (1959) introduced probabilistic methods in graph theory. The authors’ main idea can be summarized along the following lines:

> “The random graph model … assumes that we start with N vertices and connect each pair of vertices with probability p” (Barabási & Albert, 1999, p. 510).

In this subsection, we consider the case of $N = 20$.

8.1.1 Explaining p

Of $\binom{20}{2} = 190$ possible edges, we actually realize a fraction p.

Therefore, we should expect, e.g.

- None for $p = 0$
- 38 edges for $p = 0.2$
- 76 edges for $p = 0.4$
- 114 edges for $p = 0.6$
- 152 edges for $p = 0.8$
- All for $p = 1$ (fully connected)

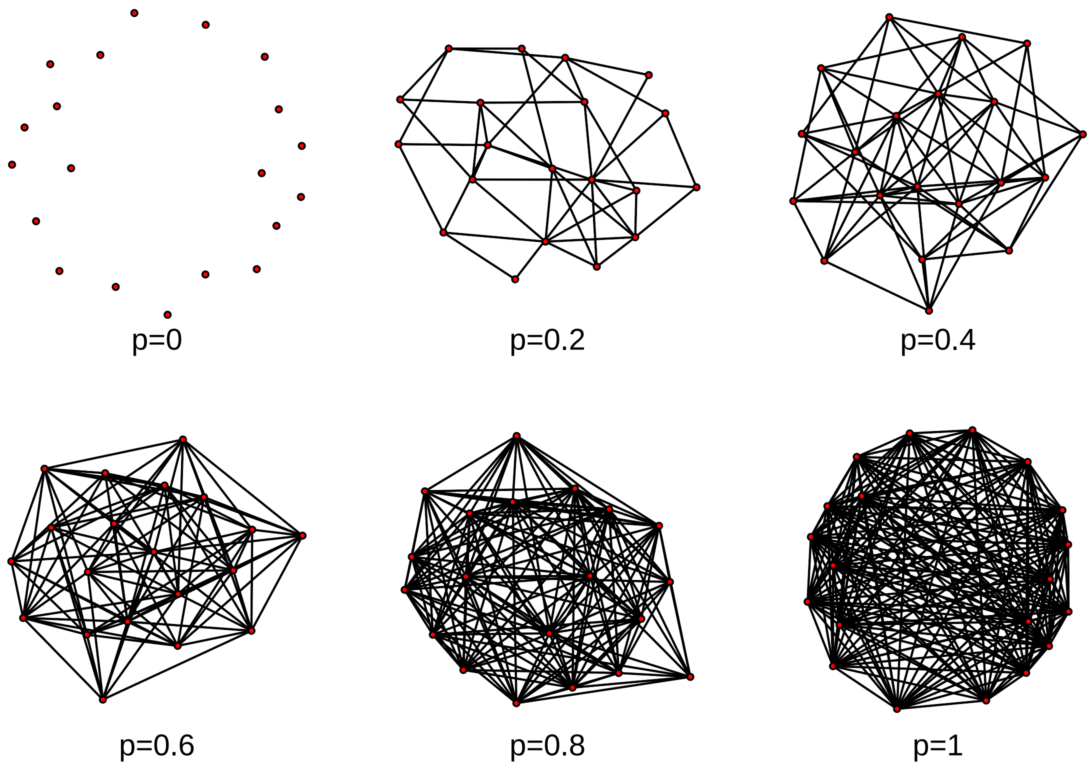


Figure 8.1: Static random networks with different values for p

Note that in Figure 8.1, we are using a combination (“Combination,” 2017) to compute the number of possible edges, which implies the following:

- Without repetition there are no *self-edges*.
- We consider *undirected* graphs, because order is not important.

8.2 Small World Model of Watts and Strogatz

Watts and Strogatz offer these instructions for creating small world networks:

> “Starting from a ring lattice with n vertices and k edges per vertex, we rewire each edge at random with probability p” (Watts & Strogatz, 1998, p. 440).

The networks depicted in Figure 8.2 derive their name from social psychological experiments from the 1960s documented by Milgram (1967). They were conducted to test whether the claim that humans inhabit “a small world” can be empirically corroborated. The folk notions of the “small world” mainly arose because many can report some apparently incredible coincidences of mutual acquaintances when meeting complete strangers.

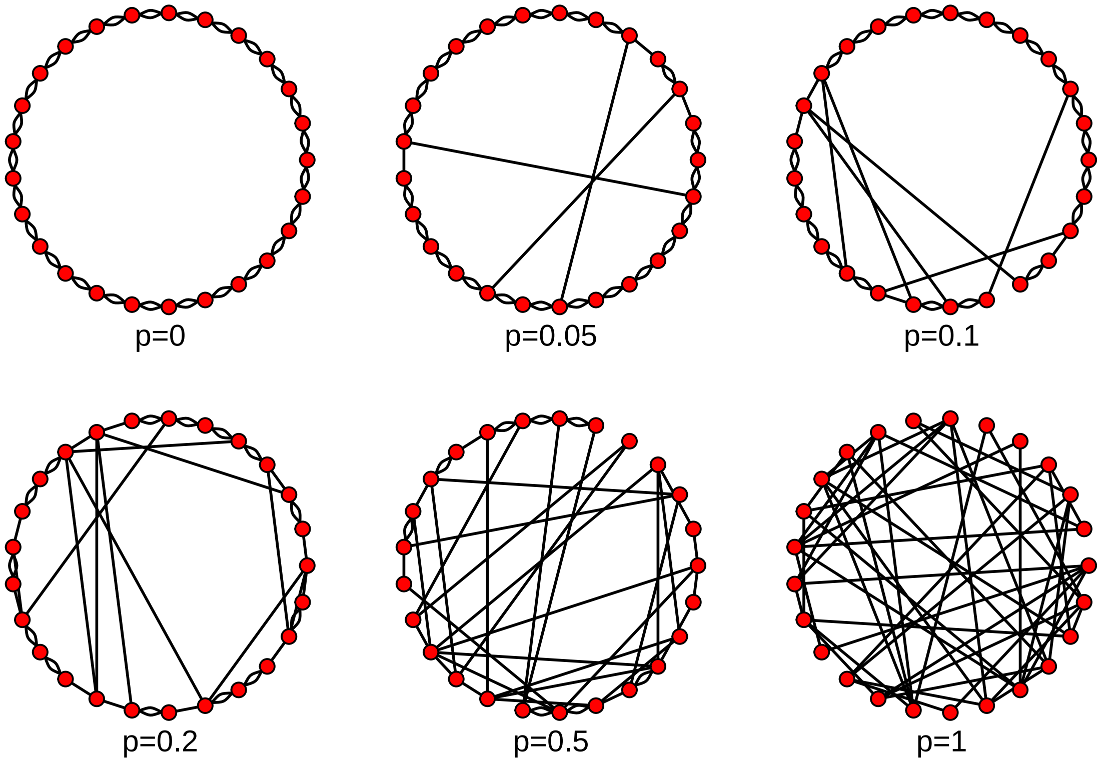


Figure 8.2: Small world networks with different rewiring probabilities p

To prepare the experiment, researchers chose two *target persons* from the vicinity of their university (Harvard) in Massachusetts on the East Coast. Then, they decided on two places in the middle of the United States: Wichita, Kansas and Omaha, Nebraska. From these two cities, they sampled several hundred *starting persons* who received a package: It contained the name of one of the target persons and instructions:

> If you do not know the target person on a personal basis, do not try to contact him directly. Instead, mail this folder ... to a personal acquaintance who is more likely than you to know the target person ... it must be someone you know on a first-name basis. (Milgram, 1967, p. 64)

Then, the proband was encouraged to send a postcard justifying his choice and providing sociodemographic information of the acquaintance to the researchers. This allowed to keep track of the packages.

What were the results? In the Nebraska group, 160 participants got the package, 44 reached their target and 116[11] died out before doing so. Those that were completed, however, reached the target person through a surprisingly low number of 2 to 10 intermediate acquaintances, with a median of 5. Therefore, the author concluded that

[11] In the original paper there is a mistake: "Of 160 chains that started in Nebraska, 44 were completed and 126 dropped out" (Milgram, 1967, p. 65). In a diagram next to the text, also 44 completed itineraries are mentioned, therefore we conclude that, very likely, 126 is the wrong number.

the claim of a "small world" can indeed be empirically justified.

Subsequently, more systematic experiments by Travers & Milgram (1969) on long distance acquaintance chains led to the development of the "small-world method" in the study of social networks.

8.3 Barabási–Albert Model

The authors provide the following description for creating a Barabási–Albert (BA) network:

> ... starting with a small number (m_0) of vertices, at every time step we add a new vertex with $m(\leq m_0)$ edges that link the new vertex to m different vertices already present in the system. To incorporate preferential attachment, we assume that the probability Π that a new vertex will be connected to vertex i depends on the connectivity k_i of that vertex, so that $\Pi(k_i) = ki/\sum_j k_j$. After t time steps, the model leads to a random network with $t + m_0$ vertices and mt edges. (Barabási & Albert, 1999, p. 511)

With these instructions, BA networks can be created from different starting configurations, which introduces a certain degree of arbitrariness. For example, it is not clear ether a vertex can link to the same vertex more than once, i.e. if *multiedges* are allowed.

8.3.1 Growth

BA networks are dynamically created, i.e. they grow by the addition of new nodes. This is different from random networks and small world networks, which are static, at least in their original formulations: Their size has to be set beforehand, while BA networks can in principle grow endlessly.

8.3.2 Preferential Attachment[12]

Preferential Attachment (PA) is also referred to as a "rich-get-richer" process: Nodes with more links have a higher probability of receiving a newly created link. The number of links a node possesses is also called the *degree* of that node. "The degree distribution $P(k)$ of a network is then defined to be the fraction of nodes in the network with degree k" ("Degree distribution," 2017). A PA process creates degree distributions that are characterized by a power law. This is intriguing, because with this, findings from

[12] Figure 15.2 in the Appendix shows the history of this idea.

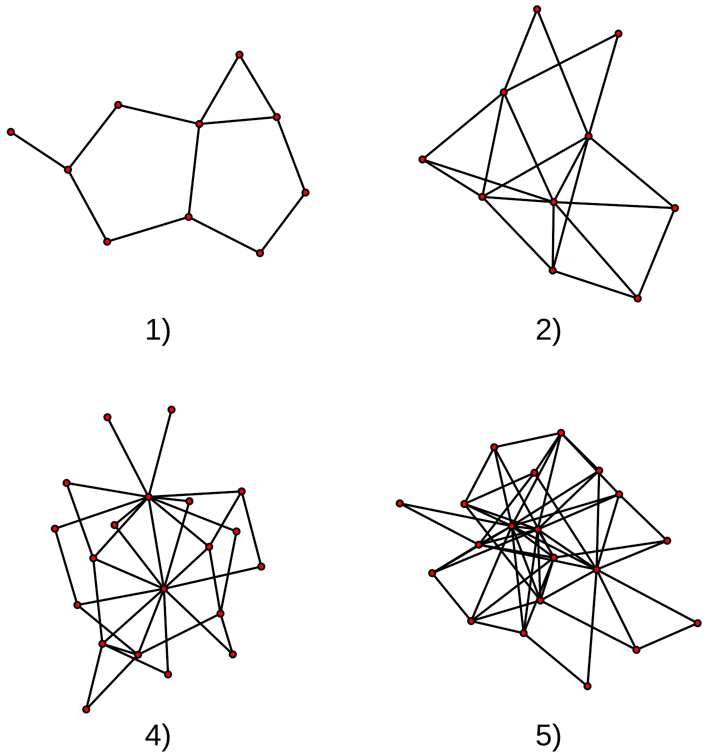


Figure 8.3: Barabási–Albert models with different starting configurations

network science are added to the wide range of empirical data where power laws are commonly found (for an early overview see Simon, 1955).

	Size	m_0	m
1)	10	2	2
2)	10	4	4
3)	10	6	6
4)	20	2	2
5)	20	4	4
6)	20	6	6

Table 8.1: Parameters used for creating the BA networks of Figure 8.3

Figure 8.4 exemplifies the dynamics of the growth of a BA network with the following parameters at different time steps:

- $m_0 = 2$
- $m = 2$
- Growing until time step $t = 30$

Vertex size corresponds to vertex degree. The formation of *hubs* (nodes with a degree that is much higher than the average degree), is clearly visible.

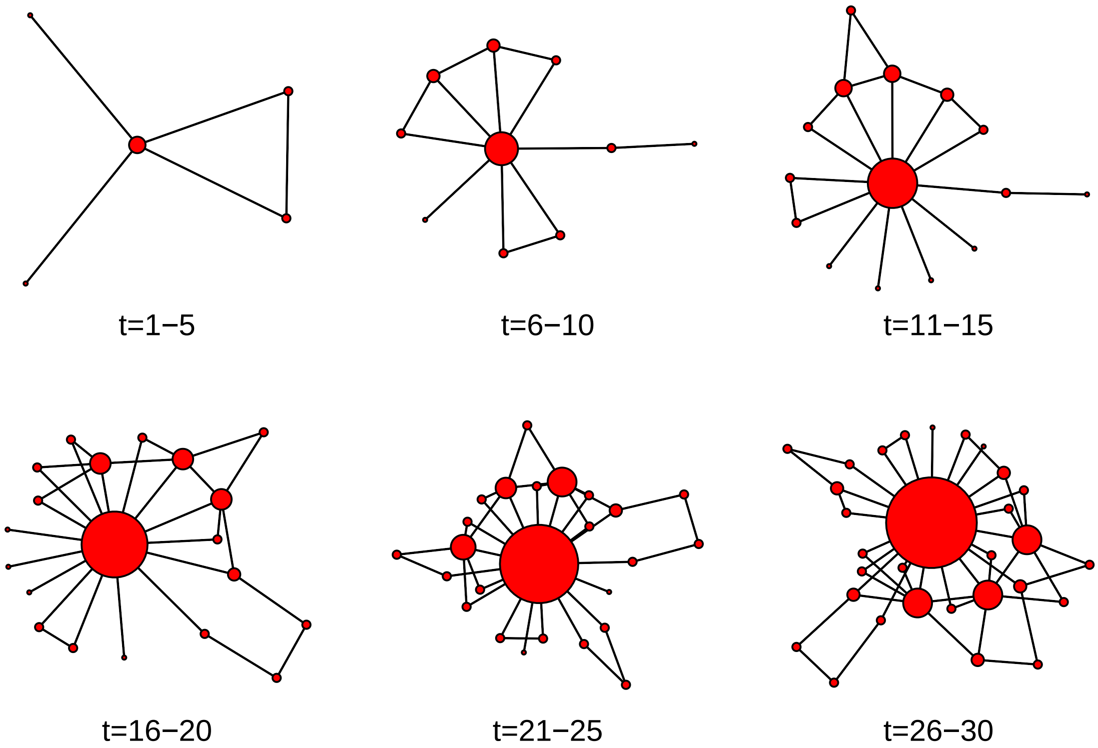


Figure 8.4: Dynamical evolution of a Barabási–Albert model

8.3.3 The Linearized Chord Diagram (LCD)

In the introduction to this Section 8.3, the original description of the process that creates a BA network is vague. For example, "as the degrees are initially zero, it is not clear how the process is supposed to get started" (Bollobás, Riordan, Spencer, & Tusnády, 2001, p. 281). Also, the precise initial configuration of m_0 is not specified. For a rigorous mathematical treatment, the need for strict protocol arose. The proposal by Bollobás et al. (2001) became the new standard, under the name "Linearized Chord Diagram (LCD)" (Barabási, 2017, Section 5.3).

As an illustration, we consider the case of $m = 1$ under LCD: We start from graph ($t = 0$) with no nodes and add one node ($t = 1$). This node creates a self-edge, as it has no other possibility than to link to itself. At $t = 2$, node 2 is created. Equation (8) gives the probabilities of $\frac{2}{3}$ of it linking to node 1 and $\frac{1}{3}$ of linking to itself. Following this schema, we can then compute the network for any number of subsequent time steps.

$$p = \begin{cases} \frac{k_i}{2t-1}, & \text{if } 1 \leq i \leq t-1 \\ \frac{1}{2t-1}, & \text{if } i = t \end{cases} \tag{8}$$

We see that, according to LCD, self-edges and multiple edges are explicitly allowed.

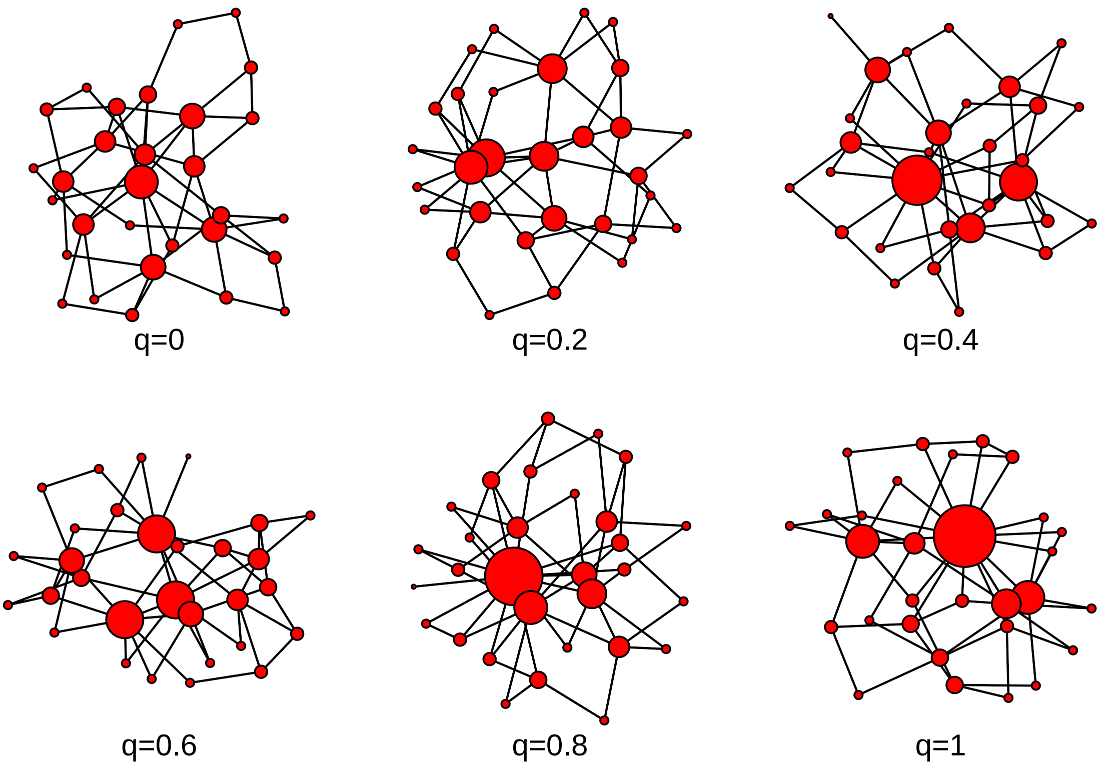


Figure 8.5: Q-networks with different values for q

8.4 q-Networks / Hybrid Networks

Namatame & Chen (2016, pp. 48–49) describe a network structure that is a hybrid between a dynamical random and a BA network, called the q-network.

8.4.1 Parameter q

A proportion q of all realized links—itself a fraction p of all logically possible links—is *preferentially attached*.

A proportion $1 - q$ is randomly attached.

- $q = 0$ creates a purely random network.
- If $q \in]0, 1[$, a hybrid is created.
- $q = 1$ creates a purely scale-free network with PA.

In Figure 8.5, vertex size again corresponds to vertex degree. The transition from a random network (upper left) to a BA network (bottom right) is characterized by the formation of hubs.

9 Collective Belief Updates

9.1 Adjacency Matrices

A network can be fully represented by an adjacency matrix. For example, this matrix shows the dynamic BA network of Section 8.3.3 (Figure 8.4) at timestep 10:

	1	2	3	4	5	6	7	8	9	10	11
1	0	1	0	0	0	0	0	0	0	0	0
2	1	0	1	1	1	1	1	0	1	1	1
3	0	1	0	0	0	1	0	0	1	0	0
4	0	1	0	0	1	0	0	0	0	0	0
5	0	1	0	1	0	0	0	0	0	0	0
6	0	1	1	0	0	0	0	0	0	1	0
7	0	1	0	0	0	0	0	1	0	0	0
8	0	0	0	0	0	0	1	0	0	0	0
9	0	1	1	0	0	0	0	0	0	0	0
10	0	1	0	0	0	1	0	0	0	0	0
11	0	1	0	0	0	0	0	0	0	0	0

Table 9.1: Adjacency Matrix

The rows show the links connecting *to* the the node specified in the row heading. The columns specify the links *from* the node specified in the column heading: 1 means there exists a link; 0 stands for "no link". (for an introduction see M. Newman, 2010, Part II). Because our network is undirected, the adjacency matrix is symmetric. If node 6 connects to node 10 ([6,10]=1), node 10 also connects to node 6 ([10,6]=1).

9.2 Influence Matrices

9.2.1 Baseline

By dividing every entry of a row by the row sum, we obtain the so-called "influence matrix" of Table 9.2. This matrix is not symmetric anymore: e.g. while node 1 is fully influenced by node 2, it is only 1 of 9 nodes influencing node 2, i.e. [2,1]=1/9. At this stage, all nodes connected to a node exert equal influence, i.e. averaging with uniform weights is applied. Also, nodes place no weight on themselves.

	1	2	3	4	5	6	7	8	9	10	11
1	0	1	0	0	0	0	0	0	0	0	0
2	1/9	0	1/9	1/9	1/9	1/9	1/9	0	1/9	1/9	1/9
3	0	1/3	0	0	0	1/3	0	0	1/3	0	0
4	0	1/2	0	0	1/2	0	0	0	0	0	0
5	0	1/2	0	1/2	0	0	0	0	0	0	0
6	0	1/3	1/3	0	0	0	0	0	0	1/3	0
7	0	1/2	0	0	0	0	0	1/2	0	0	0
8	0	0	0	0	0	0	1	0	0	0	0
9	0	1/2	1/2	0	0	0	0	0	0	0	0
10	0	1/2	0	0	0	1/2	0	0	0	0	0
11	0	1	0	0	0	0	0	0	0	0	0

Table 9.2: Influence Matrix Without Self-Edges

9.2.2 Self-influencing

In a next step, we add self-edges to account for the fact, that nodes may influence themselves, too:

	1	2	3	4	5	6	7	8	9	10	11
1	1/2	1/2	0	0	0	0	0	0	0	0	0
2	1/10	1/10	1/10	1/10	1/10	1/10	1/10	0	1/10	1/10	1/10
3	0	1/4	1/4	0	0	1/4	0	0	1/4	0	0
4	0	1/3	0	1/3	1/3	0	0	0	0	0	0
5	0	1/3	0	1/3	1/3	0	0	0	0	0	0
6	0	1/4	1/4	0	0	1/4	0	0	0	1/4	0
7	0	1/3	0	0	0	0	1/3	1/3	0	0	0
8	0	0	0	0	0	0	1/2	1/2	0	0	0
9	0	1/3	1/3	0	0	0	0	0	1/3	0	0
10	0	1/3	0	0	0	1/3	0	0	0	1/3	0
11	0	1/2	0	0	0	0	0	0	0	0	1/2

Table 9.3: Influence Matrix With Self-Edges

In this case, a node places as much weight on itself as it does on the other nodes connected to it. We can also introduce a parameter (say, s) that defines how much more weight it puts on it itself (for example, double or triple the weight it places on others). This can lead to interesting scenarios with *sticky* beliefs.

9.3 Markov Chain Theory

The self-influencing influence matrix in Section 9.2.2 is a *row stochastic* matrix, i.e. its rows add up to 1. It can be interpreted as a finite state *Markov chain*, which can be defined along these lines:

> Let the finite set of states be denoted S. If the state of the system is $s_t = s$ at time t, then there is a well-defined probability that the system will be in state $s_{t+1} = s'$ at time $t+1$. Let Π be the $n \times n$ matrix describing these transition probabilities with entries $\Pi = Pr(s_{t+1} = s' | s_t = s)$ (Jackson, 2010, p. 161).

These matrices Π are called *Markovian* to refer to their memorylessness.

9.3.1 Two Properties

It is useful to know two facts about Markov chains:

> *Irreducibility* of a Markov chain refers to the property that for any two pairs of states there is a positive probability of reaching one from the other at any future point if we start from the first.

As we are dealing with a matrix that was derived from a full representation of a network, this can be translated into saying that this network has to be *connected*. This means that “there is a path between every pair of vertices” (“Connectivity (graph theory),” 2017), i.e. there are no isolated nodes. For the BA networks and the q-networks created in this thesis this is always the case, because we always add new nodes to existing ones[13]. Small world networks are connected by definition, as they set out as ring networks. For random networks, the necessary connection probability p depends on the size n of the network. Whether a random network will be connected, can be checked with the following condition (“Erdős–Rényi model,” 2017):

> If $p > \frac{(1+\varepsilon)\ln n}{n}$ the random graph will almost surely be connected.

In the case of $n = 20$ (as in the visualizations in Section 8.1) this amounts to requiring $p \geq 0.15$ in order to ensure connectedness.

> *Aperiodicity* of a Markov chain means that “the greatest common divisor of its cycle lengths is one” (Jackson, 2010, p. 162).

The length of a cycle refers to the number of “hops” it takes from a state s to reach s again. In the examples of our networks, it suffices to say that the presence of self-edges makes the associated influence matrix aperiodic.

[13]Note that if one creates BA networks strictly according to LCD, there is the possibility (albeit a very remote one) that a node comes into existence, links to itself and receives no links from then on.

9.3.2 Steady State

The "Perron–Frobenius theorem" implies that the largest eigenvalue of an irreducible matrix with non-negative entries is real-valued (Jackson, 2010, pp. 77–78; "Perron–Frobenius theorem," 2017). For a row stochastic matrix[14] (like our influence matrix with self-edges in Section 9.2.2), this eigenvalue will be 1 (unit eigenvalue) and its associated **left-hand** eigenvector will be non-negative. If we normalize this eigenvector to 1, it then contains the so-called *steady state* distribution of the matrix.

9.3.3 Interpretation in Opinion Dynamics

It is possible to derive very precise predictions about the long run behaviour of a system that can be described as a Markov chain: If the steady state is reached, the system stays there. In our context, this means that the steady state distribution of the influence matrix (which is a Markov chain) contains the *limiting influence* of every agent on the network, i.e. the influence each agent has in the steady state.

9.4 The Collective Belief Update Mechanism

The information from the previous sections allows to detail the way the actual collective belief update works. First, we create a specific network configuration that is converted into an adjacency matrix. This is then turned into an influence matrix. Next, we obtain the limiting influences through computation of the normalized[15] eigenvector corresponding to the unit eigenvalue of the influence matrix.

In the next step, a certain number of individual belief updates is computed[16]. After the series of (typically: 10) individual belief updates—called a **phase**—one collective update takes place. In a simplified example of only five agents, this would resemble the following:

$$\begin{pmatrix} 1/5 & 2/15 & 2/15 & 1/5 & 1/3 \\ 1/5 & 2/15 & 2/15 & 1/5 & 1/3 \\ 1/5 & 2/15 & 2/15 & 1/5 & 1/3 \\ 1/5 & 2/15 & 2/15 & 1/5 & 1/3 \\ 1/5 & 2/15 & 2/15 & 1/5 & 1/3 \end{pmatrix} \begin{pmatrix} 0.00 & 0.01 & 0.06 & 0.13 & 0.19 & 0.17 & 0.24 & 0.14 & 0.05 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.02 & 0.08 & 0.15 & 0.17 & 0.28 & 0.20 & 0.07 & 0.01 & 0.00 \\ 0.00 & 0.00 & 0.01 & 0.05 & 0.11 & 0.16 & 0.31 & 0.24 & 0.10 & 0.01 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.03 & 0.08 & 0.13 & 0.31 & 0.29 & 0.14 & 0.02 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.01 & 0.05 & 0.11 & 0.30 & 0.32 & 0.18 & 0.03 & 0.00 \end{pmatrix}$$

[14] Again, this term denotes a matrix with rows that all add up to 1.
[15] Again, normalized to 1.
[16] See, Section 2.

The left matrix contains the limiting influence of the agents.

The right matrix contains:

- Rows: A subjective probability distribution for the respective agent over the eleven biases.
- Columns: The degree of belief in the respective hypothesis of every agent about each of the (possible) biases of the coin.

By multiplying the two matrices we perform a *collective belief update*. We arrive at a matrix containing the *limiting beliefs* of the agents. A new phase of individual updating is performed and is then followed by another collective update, along the lines described. This process if repeated until an established number of phases is reached.

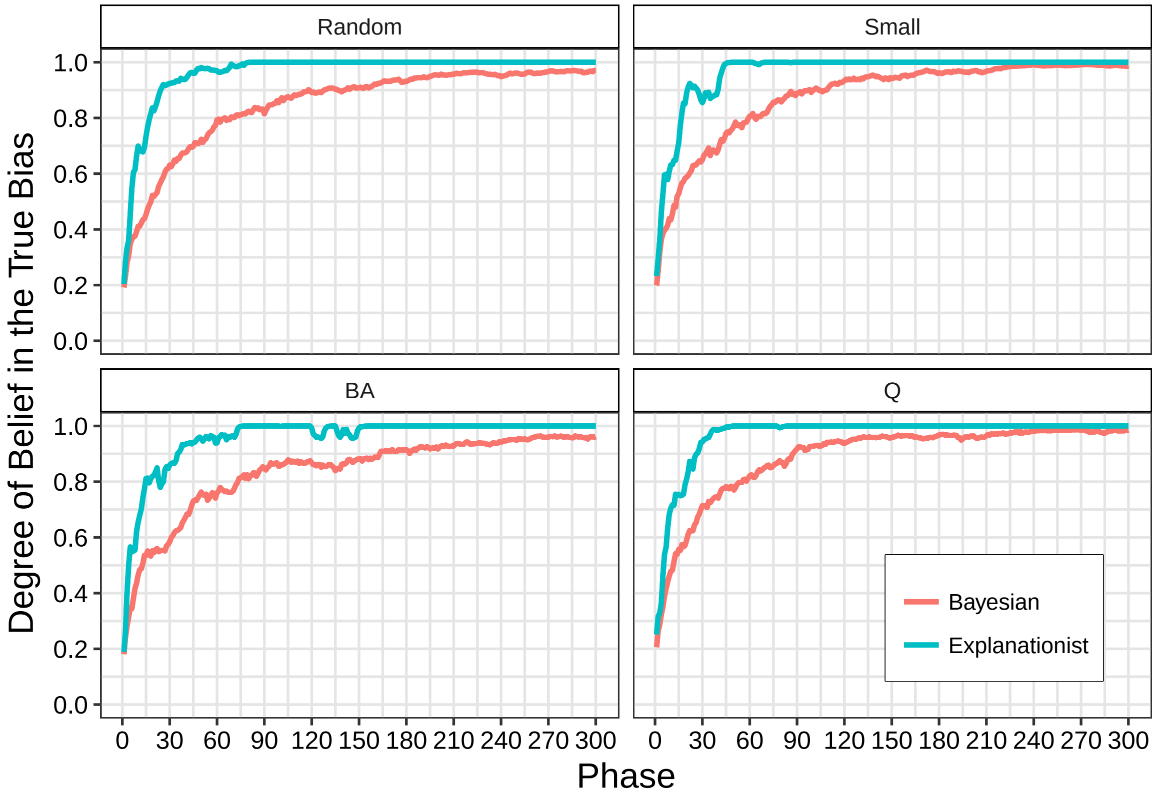
10 Comparison of network types

10.1 Certainty of Evidence

We ran simulations of 300 phases, i.e. coins are tossed 300×10 times in total, with the following settings:

- Each of the four different types of networks is populated with 50 agents of each type.
- Evidence is fully certain.
- The true bias is 0.5.

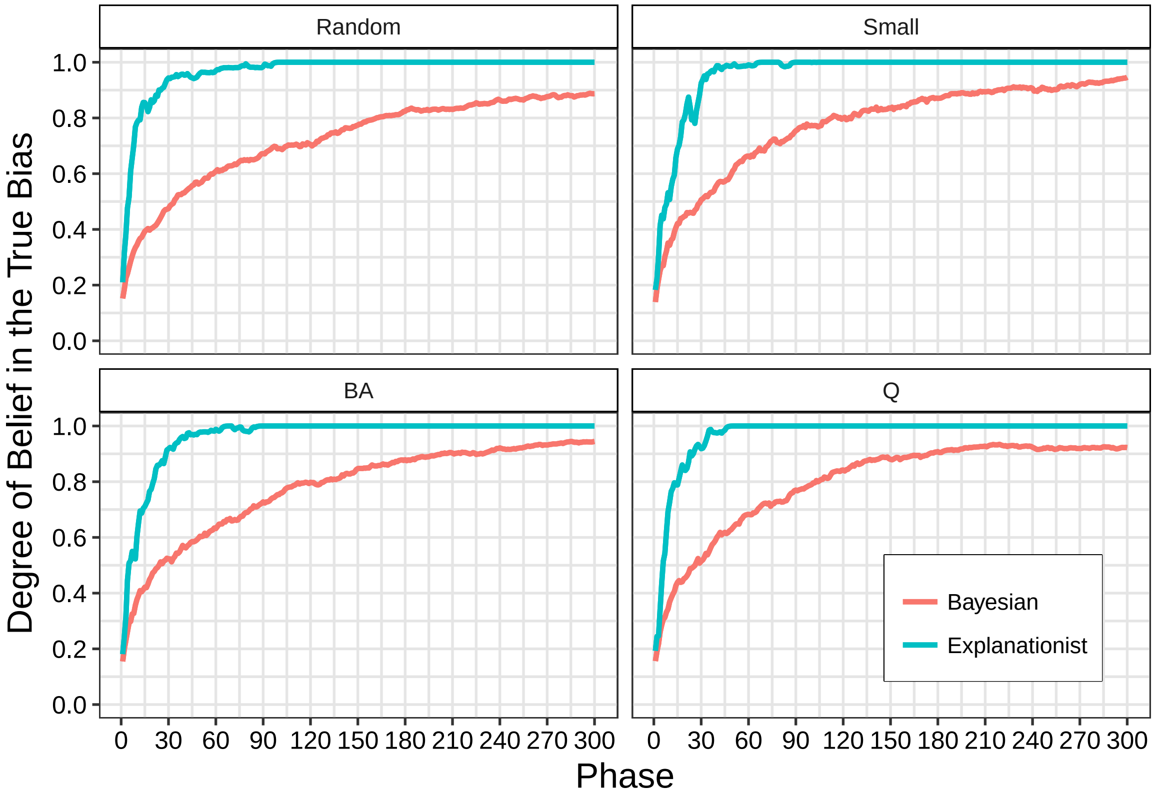
In the following you can see a comparison of the results of updating according to the two agent types on different networks:



We see that the speed advantage of the explanationist (using IBE) persists also in the collective scenario. IBE reaches a very high threshold (> 0.99) within 300 phases on all of the network topologies. The (Standard) Bayesian does so only on the small world network and q-network.

10.2 Uncertainty of Evidence

In this section, certainty of evidence is uniformly distributed over all agents with a mean of 0.79 and standard deviation of 0.12. The mean uncertainty therefore corresponds (almost) to the value used in Section 6 (“fixed certainty”). Everything else is as before.



The speed advantage of the explanationist (using, in this case, IBE*) is now more pronounced. Also, the Bayesian (JC) is not as accurate. It does not arrive at a degree of belief in the true bias that is as high as that of IBE*.

10.2.1 Findings from the Collective Domain

Generally, collective belief updating is slower than individual updating. The explanationist shows collectively even more pronounced speed advantages than in individual updating.

10.3 The Influence of Network Topology on Speed

How do the parameters used in creating networks influence the number of (individual and collective) updates needed to (permanently) cross threshold? To answer this question, we created several networks of each type with different parameters. The values reported in the tables below denote the number of phases. Minimum values are highlighted to allow for easier comparison.

The following parameters stay fixed:

- We simulate 500 phases and therefore $500 \times 10 = 5000$ tosses.

- We create 50 agents per network.
- Certainty is uniformly distributed over all agents in $[0.6, 1]$, which results in a mean of ≈ 0.8[17].
- Bias is 0.5.

Table 10.1: Random Networks: # of phases needed to cross threshold of 0.95 for different connection probabilities p

	0.25	0.3	0.35	0.4	0.5	0.6	0.7	0.8	0.9	1
Bayes	347	**337**	340	495	-	490	-	457	491	-
Expl	49	47	45	38	**37**	**37**	**37**	**37**	**37**	**37**

When p becomes larger than 0.35 for random networks, the Bayesian starts to slow down (or does not cross threshold at all). This is different for the explanationist: If the number of realized edges $\binom{N}{2} \cdot p$ increases with p, the explanationist becomes quicker.

Table 10.2: Small World Networks: # of phases needed to cross threshold of 0.95 for different rewiring probabilities p (continued below)

	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6
Bayes	299	266	**228**	229	229	246	252	403	346	342	339	416	388
Expl	44	49	47	44	47	44	44	38	36	37	36	37	40

	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1
Bayes	396	275	288	288	342	265	304	281
Expl	37	37	**33**	37	**33**	38	**33**	38

The explanationist is quicker on small world networks with high p. The performance of the Bayesian is better with small values of p.

For the analysis of BA networks, column labels refer to the number of nodes in the starting configuration (m_0). Row labels show the number of edges that are added from each node that came into existence to already existing nodes (m).

[17]For a (continuous) uniform distribution we can compute the expected mean $E(X) = \frac{1}{2}(a + b)$, with a, b being the boundaries.

	2	4	6	8	10
2	306	415	393	419	393
4	261	356	367	429	438
6	284	403	418	398	457
8	**251**	308	393	372	379
10	287	369	393	-	398

Table 10.4: BA Networks: # of phases needed to cross threshold of 0.95 for different starting configurations (Bayesian)

	2	4	6	8	10
2	38	39	34	34	39
4	38	39	**29**	34	39
6	39	39	34	**29**	39
8	38	30	34	39	39
10	38	**29**	35	39	39

Table 10.5: BA Networks: # of phases needed to cross threshold of 0.95 for different starting configurations (Explanationist)

For the following matrices on q-networks (Tables 10.6 and 10.7), column labels correspond to the number of nodes in the starting configuration *and* to the number of links added per round. Rows depict the value of q.

	2	4	6	8	10
0	243	332	399	416	352
0.1	322	323	455	455	297
0.2	398	334	454	275	364
0.3	341	339	292	331	352
0.4	340	**229**	319	457	373
0.5	295	299	405	456	352
0.6	301	373	407	375	260
0.7	340	455	417	290	375
0.8	415	472	286	370	345
0.9	341	301	340	353	443
1	241	353	407	448	357

Table 10.6: q-Networks: # of phases needed to cross threshold of 0.95 for different starting configurations (Bayesian)

	2	4	6	8	10
0	43	44	38	37	39
0.1	53	49	38	38	40
0.2	48	48	40	51	40
0.3	53	49	51	40	38
0.4	49	53	40	40	37
0.5	50	53	40	40	38
0.6	48	48	38	37	41
0.7	49	38	38	37	38
0.8	48	38	43	40	39
0.9	44	37	37	39	42
1	53	44	38	39	**32**

Table 10.7: q-Networks: # of phases needed to cross threshold of 0.95 for different starting configurations (Explanationist)

For the explanationist, a q-network with parameters that disable the creation a hybrid is the fastest: With $q = 1$, it is in fact a BA network. This is in line with the findings at the beginning of this section, which show that random networks (at the opposite extreme of q-networks) are generally slower to converge than BA networks.

In summary, it can be said that in the collective scenario the explanationist is *always faster* in assigning a degree of belief above threshold of 0.95 to the true bias. This holds for all the different network topologies examined in this section.

11 Objections and Challenges

In this section we want to provide some remarks on controversies regarding the choice of specific parameters and concepts used in our simulations.

11.1 The Threshold

It may be argued that the value for the threshold that the agents have to pass is set arbitrarily. To some extent, this is the case. Nonetheless, this admission does not mean that the problem has not been subjected to systematic inquiry. In this section we try to provide arguments for the specific threshold values researched, in order to answer the key question: Which degree of belief must be reached to make it rational for an agent to believe in a proposition?

A good first start is to choose a number other than 1, but still close to it, e.g. 0.99. We do not want to make absolute certainty our criterion, because that would be overly strict: For example, in the simulation runs documented in this thesis, 1 is generally never reached[18]. Therefore, a threshold of 0.99 sounds like a good, first proposal. However, that is not so. Let us consider the following example ("Lottery paradox," 2017): Suppose that there is a lottery of 100 tickets, with 1 winning ticket. Because P(Ticket 1 does not win)=0.99, under the chosen threshold it is rational to believe that. The same is true for P(Ticket 2 does not win)=0.99, P(Ticket 3 does not win)=0.99 and so on, until P(Ticket 100 does not win)=0.99. If you believe that every single ticket will not be the winning ticket, it is rational to believe the conjunction, i.e. no ticket will win. Of course, now you run into a contradiction: You know that the lottery is fair and believe that one ticket will win, while at the same time you believe that no ticket will win. It turns out that the threshold cannot be different from 1 under these assumptions[19]. But, we already excluded this possibility. Therefore, the so-called "lottery paradox" poses a subtantial problem.

On the other hand, a natural lower bound seems exceed a threshold of 0.5, because, at least, "you need to have more confidence in a proposition than in its negation" (Foley, 1992, p. 112). But, this time another paradox (D. C. Makinson, 1965) refutes this proposal: You write a book in which you make many claims. You can defend all these claims, i.e. your degree of belief is above the threshold (which is not 1, but sufficient for belief). It is common for an author to take responsibility for any errors in their book's preface. You do so as well, because a lengthy book (like yours) is likely to contain a few errors. Again, a contradiction occurs: You believe that all your claims in the book are correct, while at the same time you believe that at least one of them may be wrong. This contradiction even arises if you choose a threshold below 0.5: Also then, you have to believe the conjunction, which contradicts the preface.

Generally, in philosophy of science "the relationship between the rationality of beliefs and the rationality of degrees of belief" became known as "the Lockean thesis" (Foley, 1992, p. 111). There is an ongoing tradition of research in this domain, and it is still heavily debated (for an overview see Huber, 2016).

11.2 Measures of Explanatory Goodness

In the simulations of this thesis, the best hypothesis is identified with a "frequentist" measure. It is especially suited for the given scenario: If there is a binary event $\{head, tail\}$, it can be coded $\{1, 0\}$. Then you sum up the observed evidence so far and divide it by the number of tosses. You get a number in $[0, 1]$ that can "naturally" be

[18] Although, after very many belief updates on coin tosses, the degree of belief in a bias may grow that close to 1 that the internal logic of the programming language R eventually rounds it to 1.

[19] It may be possible to "sidestep" the lottery argument by giving up the conjunctivity assumption (Foley, 1992, p. 117).

translated into the "most explanatory" bias[20] so far.

Of course, this is just one proposal that is well suited for the scenario investigated. Speaking logically, there is an infinite number of possible measures: "IBE is best thought of as a slogan that can be fleshed out in different ways" (Douven, 2016, p. 1). Douven & Schupbach (2015a, p. 4) collect the following more advanced measures from the literature:

Popper's measure: $\frac{Pr(E|H) - Pr(E)}{Pr(E|H) + Pr(E)}$

Good's measure: $\ln \frac{Pr(E|H)}{Pr(E)}$

The measure of Schupbach & Sprenger (2011): $\frac{Pr(H|E) - Pr(H|\neg E)}{Pr(H|E) + Pr(H|\neg E)}$

11.3 Network Hype

As mentioned earlier, there has been a boom in network science since the end of the 1990s. In particular, networks that display scale-free properties, such as BA networks, have become a center of attention. For many decades, evidence has been collected from many different contexts, demonstrating the importance of power laws (e.g. from economics, sociology and biology, see Simon, 1955). On the one hand consilience—the convergence of evidence from (disparate) areas—adds to the relevance of the concept[21]. On the other hand, there are also more and more dissenting voices that claim that power laws are not as ubiquitous as often proclaimed (e.g. Clauset, Shalizi, & Newman, 2009).

It may be argued that a similar hype arose and, arguably, died out again around the concept of "genetic algorithms."[22] It is still too soon to say whether network science will share such a fate (or possibly undergo seasonal cycles, as the field of AI).

[20] Or into the two equal best biases, in case of a tie.

[21] For a recent treatment of the general concept, see Bookstein (2014). For a historical introduction, Snyder (2012, Section 3).

[22] In this development, the name which is hinting at "confirmed, hard science" played an important role.

12 Epistemological Considerations

12.1 Simulations

Speaking generally, we can treat as a simulation "any system that is believed, or hoped, to have dynamical behavior that is similar enough to some other system such that the former can be studied to learn about the latter" (Winsberg, 2015, Section 1.1). In this view, a simulation shares many characteristics of a model[23]: It can be studied as a *surrogate* of some interesting phenomenon.

A classical example of a simulation is the working hydraulic scale model of the San Francisco Bay and Sacramento-San Joaquin River Delta System[24], completed in 1957 ("U.S. Army Corps of Engineers Bay Model," 2017). It was build on a scale of 1:1000 on the horizontal axes and 1:100 on the vertical axis. The authorities used it to test the effect of interventions in the area that were proposed by engineers.

12.2 Computer Simulations

Computer simulations have their roots in nuclear physics and weather science. Defined narrowly, "a computer simulation is a program that is run on a computer and that uses step-by-step methods to explore the approximate behavior of a mathematical model" (Winsberg, 2015, Section 1.1) Computer simulations began to be used in physics in the analysis of systems characterized by differential equations that could not be solved analytically. The more and more widespread availability of digital computers made simulations a (potential) tool of almost any scientific discipline.

For the present purposes, we can almost work with the narrow definition above, we just have to alter one part of it: We are not only interested in *equation-based* modeling, but in *agent-based* modeling. In this context, an agent is best thought of as a container. It can be filled with any interaction rules, not exclusively mathematical ones. Crucially, we are interested in non-equilibrium states: These are often hard to model purely mathematically.

One of the main strengths of agent-based simulations lies in their ability to extend current scientific knowledge. Conventional, purely mathematical assumptions can be relaxed or replaced by local interaction rules. Therefore assumptions show that they are *heuristic*. They are stepping stones, that are employed in a method of successive approximation (Musgrave, 1981, p. 383) towards more and more realistic ones.

[23]In fact, the two names are commonly used interchangeably.
[24]Figure 15.3 shows it as it is currently displayed in a museum.

Although computer simulations share many characteristics of conventional (mathematical) models, they introduce (at least) one new component: Errors that may arise from the specific system used for creating the simulation. This includes malfunctioning of programming packages, compilers and hardware, among other things. An example would be the specific way of "randomizing the agent call order, where various methods — with different effects on output — are possible" (Epstein, 1999, p. 52).

Its pioneers saw computer science as the empirical study of machines: "Each new machine that is built is an experiment" (Newell & Simon, 1976, p. 114). At the same time, computer simulations run on these machines may constitute "experiments" of their own. It can be argued that these experiments are identical to those in the natural sciences. Classic experiments can be conducted—if not on the physical targets themselves—on a substitute that is often qualitatively similiar. The "in silico" experiments run on computers are generally more abstract. Nonetheless, this can prove to be an advantage if experiments on the study material itself cannot be conducted because they would be practically impossible or unethical, for example.

To make the attempt at developing a binding thread, a fitting characterization of the type of computer simulation developed in this thesis could be argued along the following lines:

> *A computer simulation is a program that is run on a computer and that uses step-by-step methods to explore the inter-actions between agents and the intra-actions between the constituting processes of an agent.*

12.3 The Context of Social Science

In his methodological discussion, Epstein (1999) puts forwards the argument that the adoption of agent-based modelling made the social sciences *generative*. According to Epstein, this approach cannot be classified as neither inductive or deductive in a classical sense[25]. He dismisses the first because no generative social scientist just "assembles macroeconomic data and estimates aggregate relations econometrically" (p.43). His argument stops here, but it can be extended: In general, no researcher is ever able to proceed in a naive inductive way. This is due to several reasons, one of them being the

[25] An inductive inference "passes from *singular statements* . . . such as accounts of the results of observations or experiments, to *universal statements*, such as hypotheses or theories". Deduction works the other way round: "From a new idea . . . - an anticipation, a hypothesis, a theoretical system, or what you will - conclusions are drawn by means of logical deduction" (Popper, 1959/1983, pp. 27–34). These conclusions (also called predictions) can then be tested empirically. Their veri- or falsification extends also to the theory they originate from.

problem of theory-ladenness. Leaving aside perception[26], there are at least two more aspects of this argument.

Semantically, there will never be a *neutral* term for an observation. "[T]heoretical commitments exert a strong influence on observation descriptions" (Bogen, 2017). In the history of science, a swinging stone became a pendulum for some but not for others, who stayed with seeing constrained fall. With the concept of "transaction costs" in economics, accounts of the emergence of institutions are likewise understood differently.

Background knowledge also co-determines the salient features of an observation. It may well be that different data will be collected by researchers with different theoretical beliefs. If a theory does not account for systemic risks, its proponents will not spend time researching the tipping points of large interconnected structures such as the interbanking network (research that has in fact only recently been carried out, cf. Battiston et al., 2016). Data is not given, but collected because of decisions driven by theoretical demands. For example, the demand to collect more precise information (tax records) about the distribution of wealth is driven by a newly established (or better: *revived*) research program on inequality (A., 2014).

Deduction is more closely related to generative social science. Theoretical computer science can show that in principle "for every computation there is a corresponding logical deduction" (Epstein, 1999, pp. 43–44). But, generative implies *constructive*: In deduction, this is not always the case, as non-constructive existence proofs show.

Epstein advocates to choose between models according to their power in prediction. If two models are equally good in predicting a macrophenomenon, the underlying microstructures should be compared. The one that is more in line with the data collected in laboratoy experiments should be preferred. Good agent-based models capture structures *and* dynamics.

In principle, one could come up with differential equations that describe the dynamics of a system over time. Why would one need agent-based models then? The answer lies in their explanatory power: Agent-based models are characterized by a *bottom-up* approach. It is possible to draw inferences from agent behaviour to (emergent) macro phenomena. A set of differential equations is "devoid of explanatory power despite its [potential] descriptive accuracy" (Epstein, 1999, p. 51).

All of the reasoning above culminates in the generativist motto:

[26]In this context, Thomas Kuhn invokes his famous duck-rabbit (Figure 15.1 in the Appendix). The picture "shows that two men with the same retinal impressions can see different things." This implies, that "sensory experience" is not "fixed and neutral".(1970, pp. 126–127) Although they are commonly used interchangeably—from an etymological point of view—the word *fact* (Latin "factum", "that which was made"), is more compatible with this view than the word *data* (Latin "datum", "that which was given").

> "If you didn't grow it, you didn't explain its emergence" (Epstein, 1999, p. 43).

This amounts to creating a simulated world as a means to test theories against the real world.

13 Conclusions and Outlook

13.1 Interdisciplinarity

We are using methods from computer science to investigate questions from philosophy, in particular from the subfield of formal epistemology. We are only aware of scant and scattered work[27] on computational methods in philosophy (an early, rudimentary example is Van Fraassen, Hughes, & Harman, 1986), although simulations carry huge potential for this field. Additionally, we build on, compare, and aim to reconcile and integrate work in the social sciences on agents and networks. This is complemented by mathematics (graph theory). Network science arose out of graph theory, and is by now one of the core disciplines of complexity science, itself a highly interdisciplinary field. Finally, statistics is not only at the heart of Bayesian modeling, but is also used to evaluate simulation results.

13.2 Computing Power

In the case of collective belief updates, even with (only) 50 agents, a standard laptop of 2014 (such we currently use) showed its limits. Especially the analysis of the effect of different parametrisations (as in Section 10.3) leads to long waiting times. With access to more computing power, it would be possible to devise much more complex scenarios of collective dynamics.

Memory limitations are an important class of problems that arise when using the programming language R: The exclusive storage of data structures in working memory ensures quick computations but creates frequent (overflow) errors. This is a relict of the past to some degree. By now, new performant hardware is commonly available. In Principal, it should be possible to use solid state drives as virtual memory if needed. At the moment, however, without the use of (mostly unstable) external packages, the manipulation of large data structures (for example, adjacency matrices of very large networks) is limited by the available RAM. Access to professional systems equipped

[27] Although this may be changing, see for example Knobe (2015).

with large working memory enables collection of much more fine-grained temporal data on a much bigger scale (e.g. with hundreds or thousands of agents).

13.3 Further Extensions

The next logical step for the scenarios investigated in this thesis would be an implementation of a version of the “DeGroot model” developed by Krause (Jackson, 2010, pp. 293–327). There, agents connect to a subset of other agents and update *only if* the other agents have beliefs similiar to them. This amounts to establishing an intervall of acceptable beliefs for every agent. Any of the other agents that falls into this intervall is an acceptable updating partner and gets linked. In this way, the network becomes partitioned into sets of agents that perform internal belief updates. Agent interactions play the dominant role in the model. On the downside, the mathematics becomes more complicated, e.g. consensus of beliefs is not always guaranteed. Also, only indirect control for the effect of a specific network topology is possible, because the agents create the network themselves. Possible interventions then include manipulating the range of acceptable beliefs, or the distribution of certainties over all agents (e.g., via so-called coordination media).

It could also be expected to be rewarding to work out the differences of network topologies in more detail. To do so, it may be illuminating to start with very simple “grid networks”, to have an even more stark contrast to dynamical networks.

In the end, the proof of the pudding stays in the eating: The big question is how to overcome toy model aspects. Krackhardt (1987) collected questionnaires from a small manufacturing firm and constructed an influence matrix from them. Along these lines, one could study the dynamics of information hierarchies, e.g. with the following scenario: One agent is really close to the source (almost certainty of evidence), for the other agents this varies according to their distance to the “chosen one”.

14 References

A., R. (2014, May 5). Thomas Piketty's "Capital", summarised in four paragraphs. *The Economist*. Retrieved from https://www.economist.com/blogs/economist-explains/2014/05/economist-explains

Accumulate - C++ Reference. (2017). Retrieved March 20, 2017, from http://www.cplusplus.com/reference/numeric/accumulate/

APA. (2013). Highlights of Changes From DSM-IV to DSM-5. In *Diagnostic and Statistical Manual of Mental Disorders* (Vols. 1–0). American Psychiatric Association. doi: 10.1176/appi.books.9780890425596.changes

Barabási, A.-L. (2002). *Linked: The new science of networks*. Cambridge, Mass: Perseus Pub.

Barabási, A.-L. (2017). *Network science*. Cambridge and New York: Cambridge University Press. Retrieved from http://barabasi.com/networksciencebook/

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*(5439), 509–512. doi: 10.1126/science.286.5439.509

Battiston, S., Farmer, J. D., Flache, A., Garlaschelli, D., Haldane, A. G., Heesterbeek, H., ... Scheffer, M. (2016). Complexity theory and financial regulation. *Science*, *351*(6275), 818–819. doi: 10.1126/science.aad0299

Bogen, J. (2017). Theory and Observation in Science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/sum2017/entries/science-theory-observation/

Bollobás, B., Riordan, O., Spencer, J., & Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, *18*(3), 279–290. doi: 10.1002/rsa.1009

Bookstein, F. L. (2014). *Measuring and Reasoning: Numerical Inference in the Sciences*. Cambridge University Press.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3. doi: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Brier score. (2017, March 7). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Brier_score&oldid=769136678

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-Law Distributions in

Empirical Data. *SIAM Review*, *51*(4), 661–703. doi: 10.1137/070710111

Combination. (2017, August 23). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Combination&oldid=796833438

Common Higher Order Functions in R package "base" version 3.5. (2017). Retrieved March 19, 2017, from https://stat.ethz.ch/R-manual/R-devel/library/base/html/funprog.html

Connectivity (graph theory). (2017, August 12). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Connectivity_(graph_theory)&oldid=795130614

Černis, E., Vassos, E., Brébion, G., McKenna, P., Murray, R., David, A., & MacCabe, J. (2015). Schizophrenia patients with high intelligence: A clinically distinct sub-type of schizophrenia? *European Psychiatry*, *30*(5), 628–632. doi: 10.1016/j.eurpsy.2015.02.007

Degree distribution. (2017, March 15). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Degree_distribution&oldid=770465378

Diaconis, P., & Zabell, S. L. (1982). Updating Subjective Probability. *Journal of the American Statistical Association*, *77*(380), 822–830. doi: 10.2307/2287313

Douven, I. (2013). Inference to the Best Explanation, Dutch Books, and Inaccuracy Minimisation. *The Philosophical Quarterly*, *63*(252), 428–444. doi: 10.1111/1467-9213.12032

Douven, I. (2016). Inference to the Best Explanation: What Is It? And Why Should We Care? In T. Poston & K. McCain (Eds.), *Best Explanations: New Essays on Inference to the Best Explanation* (pp. 1–18). Oxford University Press.

Douven, I., & Romeijn, J.-W. (2011). A New Resolution of the Judy Benjamin Problem. *Mind*, *120*(479), 637–670. doi: 10.1093/mind/fzr051

Douven, I., & Schupbach, J. N. (2015a). Probabilistic alternatives to Bayesianism: The case of explanationism. *Frontiers in Psychology*, *6*. doi: 10.3389/fpsyg.2015.00459

Douven, I., & Schupbach, J. N. (2015b). The role of explanatory considerations in updating. *Cognition*, *142*, 299–311. doi: 10.1016/j.cognition.2015.04.017

Douven, I., & Wenmackers, S. (2015). Inference to the Best Explanation versus Bayes's Rule in a Social Setting. *The British Journal for the Philosophy of Science*. doi: 10.1093/bjps/axv025

Epstein, J. M. (1999). Agent-based computational models and generative social science.

Complexity, *4*(5), 41–60. doi: 10.1002/(SICI)1099-0526(199905/06)4:5<41::AID-CPLX9>3.0.CO;2-F

Erdős, P., & Rényi, A. (1959). On Random Graphs I. *Publicationes Mathematicae (Debrecen)*, *6*, 290–297.

Erdős–Rényi model. (2017, July 30). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Erd%C5%91s%E2%80%93R%C3%A9nyi_model&oldid=793145730

Fein, L. (1959). The role of the University in computers, data processing, and related fields. *Communications of the ACM*, *2*(9), 7–14. doi: 10.1145/368424.368427

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*(1), 48–58. doi: 10.1038/nrn2536

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1-2), 3–71. doi: 10.1016/0010-0277(88)90031-5

Foley, R. (1992). The Epistemology of Belief and the Epistemology of Degrees of Belief. *American Philosophical Quarterly*, *29*(2), 111–124.

Frith, C. D. (1979). Consciousness, information processing and schizophrenia. *The British Journal of Psychiatry*, *134*(3), 225–235. doi: 10.1192/bjp.134.3.225

Harman, G. H. (1965). The Inference to the Best Explanation. *The Philosophical Review*, *74*(1), 88. doi: 10.2307/2183532

Hemsley, D. R., & Garety, P. A. (1986). The formation of maintenance of delusions: A Bayesian analysis. *The British Journal of Psychiatry*, *149*(1), 51–56. doi: 10.1192/bjp.149.1.51

Huber, F. (2016). Formal Representations of Belief. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/spr2016/entries/formal-belief/

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton: Princeton University Press.

Joyce, J. M. (2011). The Development of Subjective Bayesianism. In S. H. Dov M. Gabbay & J. Woods (Eds.), *Handbook of the History of Logic* (Vol. 10, pp. 415–475). North-Holland.

Kaplan, C. M., Saha, D., Molina, J. L., Hockeimer, W. D., Postell, E. M., Apud, J. A., … Tan, H. Y. (2016). Estimating changing contexts in schizophrenia. *Brain*, *139*(7), 2082–2095. doi: 10.1093/brain/aww095

Kemp, R., Chua, S., McKenna, P., & David, A. (1997). Reasoning and delusions. *The British Journal of Psychiatry*, *170*(5), 398–405. doi: 10.1192/bjp.170.5.398

Knobe, J. (2015). Philosophers are doing something different now: Quantitative data. *Cognition*, *135*, 36–38. doi: 10.1016/j.cognition.2014.11.011

Krackhardt, D. (1987). Cognitive social structures. *Social Networks*, *9*(2), 109–134. doi: 10.1016/0378-8733(87)90009-8

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions.* (O. Neurath, Ed.) (2nd ed., enlarged, Vol. 2). Chicago, Ill: Chicago Univ. Press.

Leitgeb, H., & Pettigrew, R. (2010). An Objective Justification of Bayesianism I: Measuring Inaccuracy*. *Philosophy of Science*, *77*(2), 201–235. doi: 10.1086/651317

List of cognitive biases. (2017, October 2). In *Wikipedia.* Retrieved from https://en.wikipedia.org/w/index.php?title=List_of_cognitive_biases&oldid=803430523

Lottery paradox. (2017, October 19). In *Wikipedia.* Retrieved from https://en.wikipedia.org/w/index.php?title=Lottery_paradox&oldid=806034462

Makinson, D. (2003). Bridges between Classical and Nonmonotonic Logic. *Logic Journal of IGPL*, *11*(1), 69–96. doi: 10.1093/jigpal/11.1.69

Makinson, D. C. (1965). The paradox of the preface. *Analysis*, *25*(6), 205–207. doi: 10.1093/analys/25.6.205

McCarthy, J. (1978). History of LISP. *ACM SIGPLAN Notices*, *13*(8), 217–223. doi: 10.1145/960118.808387

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*, *27*(4), 12. (Original work published 1955)

McDermott, D. (1976). Artificial intelligence meets natural stupidity. *ACM SIGART Bulletin*, (57), 4–9. doi: 10.1145/1045339.1045340

Milgram, S. (1967). The Small-World Problem. *Psychology Today*, *1*(1), 61–67.

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144. doi: https://doi.org/10.1016/S1364-6613(03)00029-9

Musgrave, A. (1981). 'Unreal Assumptions' in Economic Theory: The F-Twist Untwisted. *Kyklos*, *34*(3), 377–387.

Namatame, A., & Chen, S.-H. (2016). *Agent-Based Modeling and Network Dynamics.*

Oxford, New York: Oxford University Press.

Newell, A., & Simon, H. A. (1963). GPS, A Program that Simulates Human Thought. In *Computers and Thought* (pp. 279–296). New York and San Francisco and Toronto and London: McGraw-Hill Book Company.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, *19*(3), 113–126. doi: 10.1145/360018.360022

Newman, M. (2010). *Networks: An Introduction.* New York, NY, USA: Oxford University Press, Inc.

Oaksford, M., & Chater, N. (1991). Against Logicist Cognitive Science. *Mind & Language*, *6*(1), 1–38. doi: 10.1111/j.1468-0017.1991.tb00173.x

Pellert, M., & Trpin, B. (2017). Inference to the Best Explanation in Uncertain Evidential Situations. *Manuscript Submitted for Publication.*

Pelletier, F. J., & Elio, R. (1997). What Should Default Reasoning be, by Default? *Computational Intelligence*, *13*(2), 165–187. doi: 10.1111/0824-7935.00037

Perron–Frobenius theorem. (2017, October 2). In *Wikipedia.* Retrieved from https://en.wikipedia.org/w/index.php?title=Perron%E2%80%93Frobenius_theorem&oldid=803436717

Popper, K. R. (1983). *The Logic of Scientific Discovery* (11th impression). London: Hutchinson. (Original work published 1959)

Schupbach, J. N., & Sprenger, J. (2011). The Logic of Explanatory Power. *Philosophy of Science*, *78*(1), 105–127. doi: 10.1086/658111

Simon, H. A. (1955). ON A CLASS OF SKEW DISTRIBUTION FUNCTIONS. *Biometrika*, *42*(3-4), 425–440. doi: 10.1093/biomet/42.3-4.425

Snyder, L. J. (2012). William Whewell. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2012). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2012/entries/whewell/

Strasser, C., & Antonelli, G. A. (2016). Non-monotonic Logic. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2016/entries/logic-nonmonotonic/

Talbott, W. (2016). Bayesian Epistemology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2016/entries/

epistemology-bayesian/

Travers, J., & Milgram, S. (1969). An Experimental Study of the Small World Problem. *Sociometry*, *32*(4), 425. doi: 10.2307/2786545

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. doi: 10.1126/science.185.4157.1124

U.S. Army Corps of Engineers Bay Model. (2017, June 11). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=U.S._Army_Corps_of_Engineers_Bay_Model&oldid=785084834

Van Fraassen, B. C. (1989). *Laws and symmetry*. Oxford; New York: Oxford University Press.

Van Fraassen, B. C., Hughes, R. I. G., & Harman, G. (1986). A Problem for Relative Information Minimizers, Continued. *The British Journal for the Philosophy of Science*, *37*(4), 453–463. Retrieved from http://www.jstor.org.uaccess.univie.ac.at/stable/687370

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, *393*(6684), 440–442. doi: 10.1038/30918

Winsberg, E. (2015). Computer Simulations in Science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2015). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/sum2015/entries/simulations-science/

14.1 List of Figures

14.2 List of Tables

15 Appendix

15.1 Annotated Bibliography of Core Literature

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. Science, 286(5439), 509–512. https://doi.org/10.1126/science.286.5439.509
The authors introduce new kinds of networks, characterized by growth and preferential attachment. The degree distributions of BA networks follow a power law.

Bookstein, F. L. (2014). Measuring and Reasoning: Numerical Inference in the Sciences. Cambridge University Press.
Bookstein shows the role of two mechanisms in scientific explanation, abduction and consilience. Additionally, a host of statistical methods for scientific problem solving is presented.

Douven, I. (2013). Inference to the Best Explanation, Dutch Books, and Inaccuracy Minimisation. The Philosophical Quarterly, 63(252), 428–444. https://doi.org/10.1111/1467-9213. 12032
The blueprint for the simulated updating scenario. It is shown that Inference to the Best Explanation (IBE) is generally faster, but incurs on average higher penalties for inaccuracy.

Epstein, J. M. (1999). Agent-based computational models and generative social science. Complexity, 4(5), 41–60. https://doi.org/10.1002/(SICI)1099-0526(199905/06)4:5<41::AID- CPLX9>3.0.CO;2-F
The agent-based model is the key tool to make social science truly *generative*. It is well suited for the analysis of "spatially distributed systems of heterogeneous autonomous actors with bounded information and computing capacity".

Erdős, P., & Rényi, A. (1959). On Random Graphs I. Publicationes Mathematicae (Debrecen), 6, 290–297.
One of a number of papers by the same authors introducing probabilistic ideas in graph theory. It contributed to a significant part of the foundations of modern network science.

Harman, G. H. (1965). The Inference to the Best Explanation. The Philosophical Review, 74(1), 88. https://doi.org/10.2307/2183532
This article is the first to introduce "Inference to the Best Explanation" (IBE), an idea similar to that of abduction. It is argued that induction is best viewed as a special case of IBE.

Jackson, M. O. (2010). Social and Economic Networks. Princeton: Princeton University Press.
This book provides a rich resource for researchers interested in the domain. Excellent introductory chapters provide the knowledge to cover advanced material on specific models.

Joyce, J. M. (2011). The Development of Subjective Bayesianism. In S. H. Dov M. Gabbay & J. Woods (Eds.), Handbook of the History of Logic (Vol. 10, pp. 415–475). North-Holland.
Serves as an introduction to the history of Bayesianism. Necessary mathematics are presented in an accessible way.

Lipton, P. (2004). Inference to the best explanation (2nd ed). London; New York: Routledge/ Taylor and Francis Group.
Although Lipton is critical of the idea, this work provides a good introduction to the topic of "Inference to the Best Explanation".

McDermott, D. (1976). Artificial intelligence meets natural stupidity. ACM SIGART Bulletin, (57), 4–9. https://doi.org/10.1145/1045339.1045340
A classical critique of Classical AI and Cognitive Science. It is written from an insider perspective.

Namatame, A., & Chen, S.-H. (2016). Agent-Based Modeling and Network Dynamics. Oxford, New York: Oxford University Press.
The author's main motivation is to make to *explicit* the role of networks in agent-based modelling. The material is, in most parts, presented in an advanced, yet accessible way.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. Communications of the ACM, 19(3), 113–126. https://doi.org/10.1145/360018.360022
Classical article describing the main paradigm of Classical AI (and, in turn, Classical Cognitive Science), the "Physical Symbol System Hypothesis". The authors argue for the necessity and sufficency of a physical symbol system for General Artificial Intelligence.

Newman, M. (2010). Networks: An Introduction. New York, NY, USA: Oxford University Press, Inc.
This book serves as the de-facto textbook on network theory. Newman manages to provide motivation and the basics for the systematic study of networks, while at the same time going in-depth.

Simon, H. A. (1955). ON A CLASS OF SKEW DISTRIBUTION FUNCTIONS. Biometrika, 42(3-4), 425–440. https://doi.org/10.1093/biomet/42.3-4.425
Simon discusses empirical data from a range of scientifc disciplines that show striking similarities. Today, the underlying probability distributions are called power laws.

Van Fraassen, B. C. (1989). Laws and symmetry. Oxford; New York: Oxford University Press.
The "Alien Die Model" is first introduced. Bayesian belief updating is the sole approach that is rational, as alternatives are prone to being exposed to "Dutch books".

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. Nature, 393(6684), 440–442. https://doi.org/10.1038/30918
Small world networks are introduced as a way to capture empirical regularities that cannot be replicated in random networks. One of the seminal papers that kick-started modern network science.

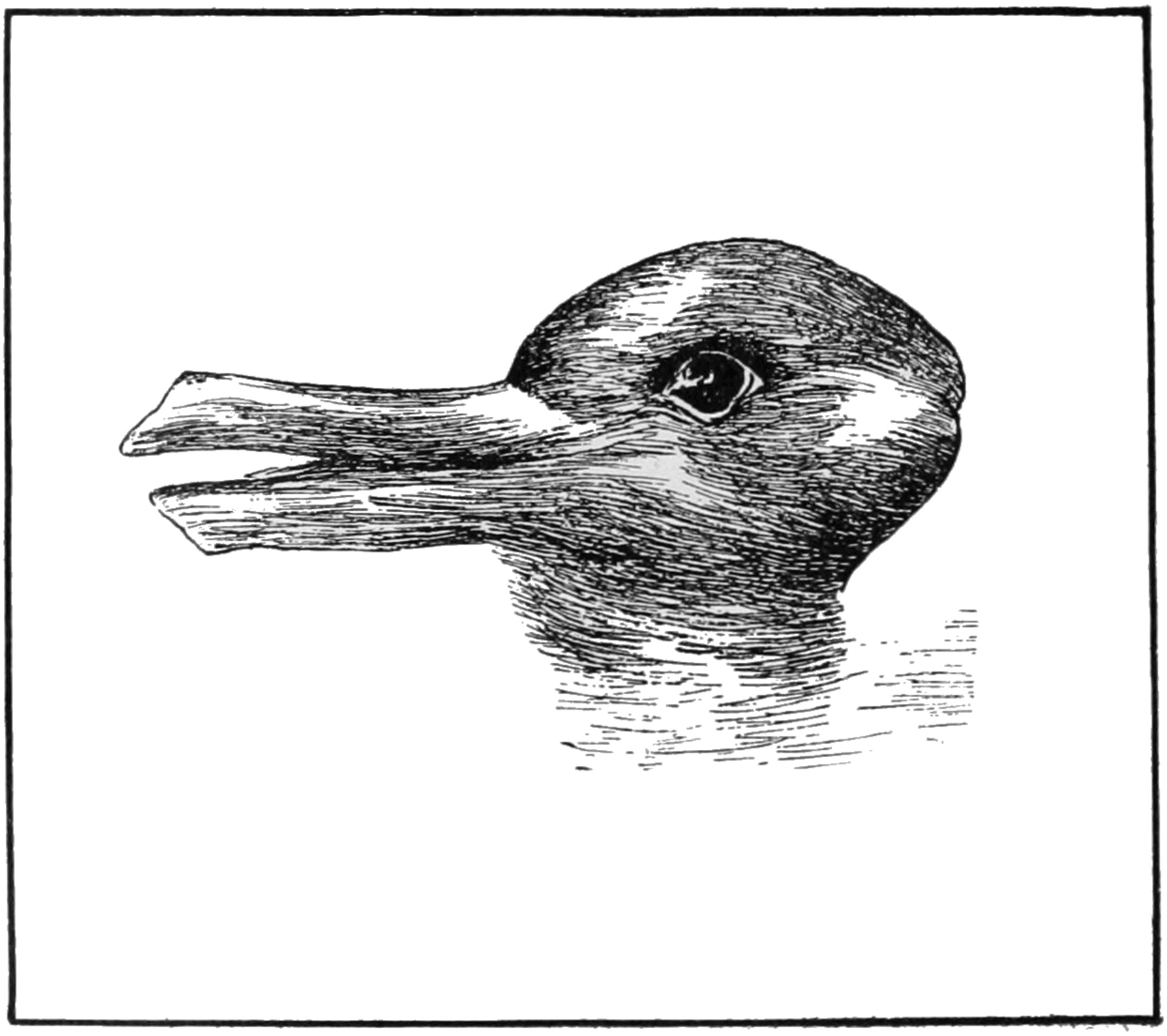


Figure 15.1: Kuhn's Duck-Rabbit

Source: https://commons.wikimedia.org/wiki/File:PSM_V54_D328_Optical_illusion_of_a_duck_or_a_rabbit_head.png

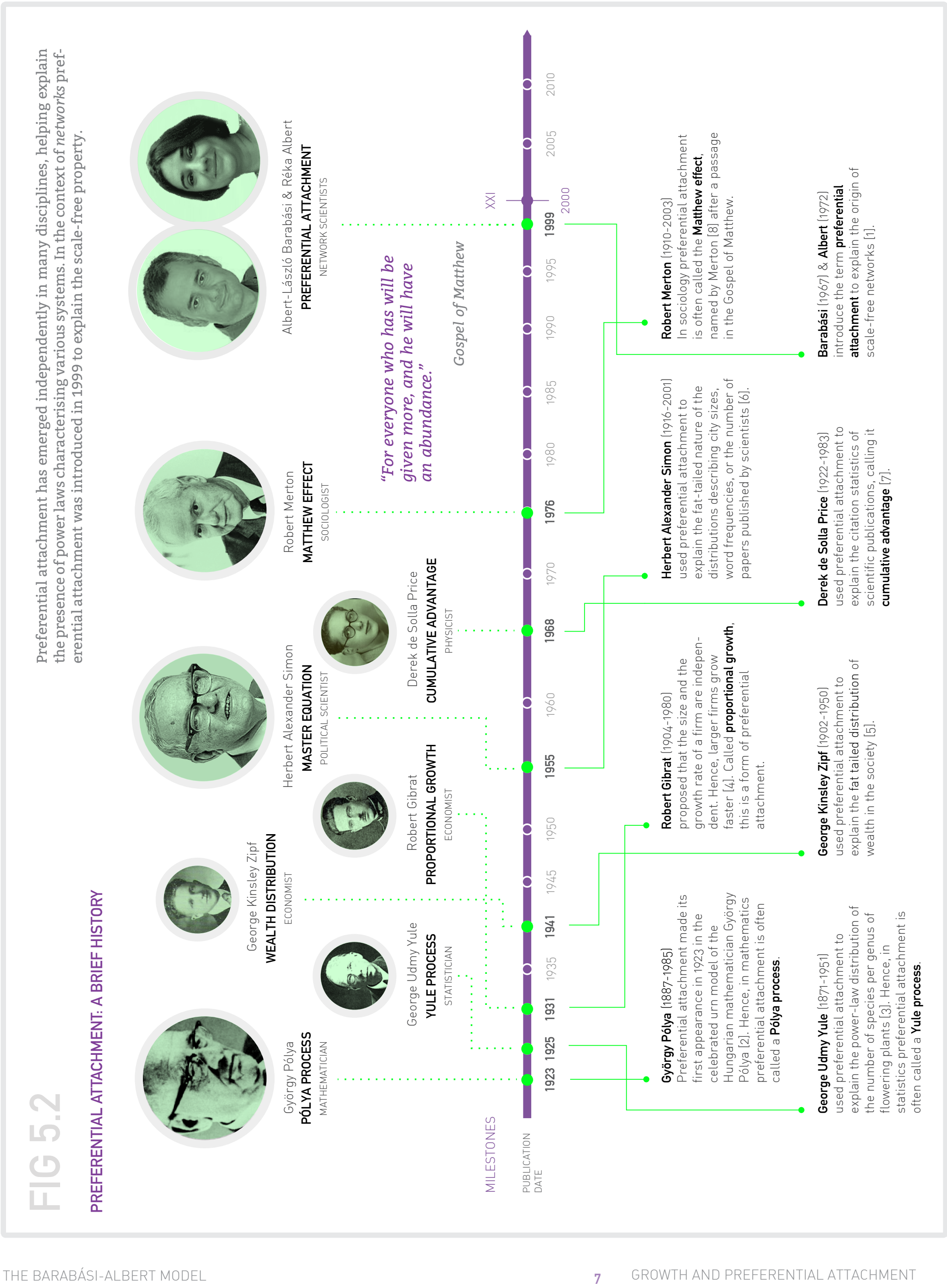


Figure 15.2: History of Preferential Attachment

Source: http://barabasi.com/f/622.pdf

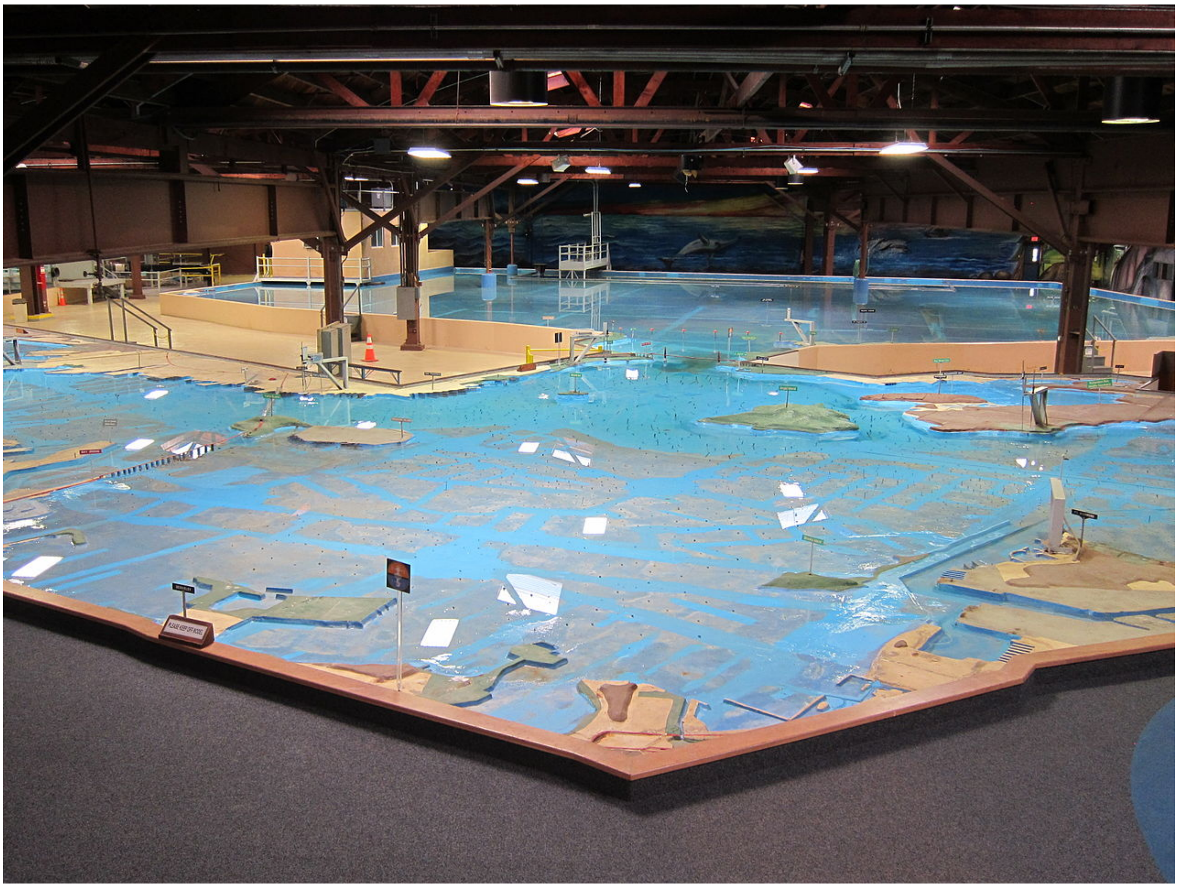


Figure 15.3: View of the San Francisco Bay portion of the Bay Model

Source: https://upload.wikimedia.org/wikipedia/commons/thumb/8/87/USCAE_Bay_Model_-_San_Francisco_Bay_Detail.jpg/1280px-USCAE_Bay_Model_-_San_Francisco_Bay_Detail.jpg

15.2 CV

Personal Data

Max Pellert, Bakk. rer. soc. oek.
1991/1/19
Schönbrunner Schlossstr. 16/12
1120 Vienna, Austria (EU)
Tel: +43 680 40 31 255
Email: a1007093@univie.ac.at

Education

10/2015 - Present	Middle European interdisciplinary master programme in Cognitive Science (MEi:CogSci), University of Vienna, Medical University of Vienna, University of Ljubljana **cand. M.Sc. (Master of Science)** (expected by 11/2017) "Collective Dynamics of Multi-Agent Networks: Simulation Studies in Probabilistic Reasoning" Master Thesis, supervised by Univ.-Lektor Dipl.-Ing. Dr. Paolo Petta
10/2010 - 10/2014	Bakkalaureat Volkswirtschaftslehre (Bachelor in Economics) (University of Vienna) **Bakk. rer. soc. oek. (Bachelor of Economics)** "Some Remarks on the Genesis of the Neoclassical Paradigm In Economics" Bachelor Thesis I, supervised by ao. Univ.-Prof. Mag. Dr. Reinhard Pirker (Vienna University of Economics and Business) "Popper and McCloskey. Logic or Rhetoric of Science?" Bachelor Thesis II, supervised by ao. Univ.-Prof. Mag. Dr. Karl Milford (University of Vienna)
06/2009	Matura mit Auszeichnung (secondary school leaving certificate with honors)

Additional Qualifications

2014 - 2015	Master study in History and Philosophy of Science (HPS), University of Vienna
2011 - 2015	Bachelor study in Psychology, University of Vienna

Languages

Natural	German: Mother Tongue English: Proficient user (C2) Spanish: Independent user (B2) French: Independent user (B1)
Programming	R: Expert Python: Advanced NetLogo: Advanced C++: Advanced

Selected Work Experience

06/2017 - Present	Study Coordinator at Hanusch Hospital Vienna (Angiologische Abteilung/Angiological Ward)

Academic Exchange Programs

09/2016 - 02/2017	ERASMUS mobility semester at the University of Ljubljana, Slovenia Collaboration with Borut Trpin (University of Ljubljana, Faculty of Arts, Department of Philosophy) on simulations in formal epistemology

Conferences

Budapest, 2017 June 22 - 24	MEi:CogSci Conference 2017 (Pellert, 2017), presentation of the results of a 10 ECTS study project http://www.univie.ac.at/meicogsci/php/ocs/index.php /meicog/meicog2017/
Bochum, 2017 June 6 - 9	Summer School on Computational Methods and Simulation for Economics Hands-on workshop with Marco Valente (University of L'Aquila)

	and Tommaso Ciarli (University of Sussex) on the "Laboratory for Simulation Development" (LSD, www.labsimdev.org/) http://www.ruhr-uni-bochum.de/mak/lehre/css.html.en
Munich, 2017 January 25 - 26	Workshop: Inferentialism, Bayesianism, and Scientific Explanation Invited Talk: "Inference to the Best Explanation in Cases of Uncertain Evidence" http://www.mcmp.philosophie.uni-muenchen.de/events/workshops/container/inf_bay_scie-workshop/index.html
Vienna, 2016 June 23 - 25	MEi:CogSci Conference 2016 (Pellert, 2016), poster about the results of a 10 ECTS study project http://www.univie.ac.at/meicogsci/php/ocs/index.php/meicog/meicog2016/
Trento, 2016 June 2 - 5	INET Workshop at Festival dell'Economia Seminar lectures by Barry Eichengreen (University of California, Berkeley) http://2016.festivaleconomia.eu/
Günne, 2016 March 4 - 11	Interdisciplinary College Week-long, intense spring school on AI, Cognitive Science and beyond http://www.interdisciplinary-college.de/
Trento, 2015 May 29 - June 2	INET Wokshop at Festival dell'Economia Seminar lectures by Joseph Stiglitz (Columbia University), Martin Guzman (Columbia University) and Steve Fazzari (Washington University of St. Louis) http://2015.festivaleconomia.eu/
Berlin, 2014 August 8 - 10	IMK-Workshop "Pluralismus in der Ökonomik" Invited Talk: "Popper und McCloskey: Konsequenzen der Kritik" http://www.boeckler.de/veranstaltung_imk_45881.htm

Publications

Pellert, M. & Trpin, B. (2017). Inference to the Best Explanation in Uncertain Evidential Situations. Manuscript submitted for publication.

Pellert, M. (2017). Collective Dynamics of Multi-Agent Networks: Simulation Studies in Probabilistic Reasoning. In P. Hochenauer,

C. Schreiber, K. Rötzer, & E. Zimmermann (Eds.), *Proceedings of the MEi: CogSci Conference 2017 in Budapest, Hungary* (June 22 - 24). Bratislava, Slovakia: Comenius University.

Pellert, M. (2016). Are Heterogeneous Expectations a Viable Alternative to Rational Expectations in Economics? In S. Ersoy, P. Hochenauer, & C. Schreiber (Eds.), *Proceedings of the MEi: CogSci Conference 2016 in Vienna, Austria* (June 23 - 25). Bratislava, Slovakia: Comenius University.