Universität St.Gallen

University of St.Gallen School of Management, Economics, Law, Social Sciences and International Affairs

# Bachelor Thesis

in the degree program
Bachelor of Economics

# Conditions for Rational Consensus: A Bayesian Perspective

**Noé Kuhn**
Matr. Nr.: 20-613-154
noeromeo.kuhn@student.unisg.ch
+41 76 512 03 20
Höhenweg 13, 5616 Meisterschwanden

on March 9, 2024
at the University of St. Gallen

*Supervisor:* Prof. Dr. Martin Kolmar

# Abstract

This thesis explores the conditions under which rational consensus can be achieved from a Bayesian perspective. Nielsen & Stewart (2018, pg. 1) correspondingly put forth *the optimistic thesis about learning* (TOTAL), which states: "Rational agents who learn the same evidence resolve disagreements". It is optimistic because this thesis does not necessarily hold. First, we provide a comprehensive epistemic account of the rational agent sufficient for TOTAL. Considering two agents that initially disagree, multiple merging theorems in Bayesian Confirmation Theory are then presented which demonstrate an asymptotic resolution of disagreement. An emphasis is put on accessibility as well as the requirements demanded for the theorems to hold. Our own detailed proofs are provided for the conjugate prior-likelihood setting. Then, we compare the theorems to highlight multiple epistemic assumptions that can be omitted as Bayesian Confirmation Theory has developed. Finally, we present a novel extension to the classic Bayesian model which may tackle one key assumption of merging theorems, namely that agents are equally dogmatic. Given the urgency for agreement on political, diplomatic, and environmental concerns, there must be awareness that even for idealistic and rational agents which learn from the same evidence, consequent agreement is less of a fortress than it is a sandcastle.

# Table of Contents

**Table of Contents** **II**

**1 Introduction** **1**

1.1 Historical Development of Bayesian Confirmation Theory . . . . . . . 3

**2 Technical Preliminaries** **5**

2.1 Set, Probability, and Measure Theory Preliminaries . . . . . . . . . . 5

2.2 The Bayesian Machinery . . . . . . . . . . . . . . . . . . . . . . . . . 7

**3 An Epistemic Agent** **10**

3.1 Bayesian Agents as Propositional, Logical, Rational . . . . . . . . . . 11

3.2 Immediate agreement . . . . . . . . . . . . . . . . . . . . . . . . . . 13

3.3 Iterative evidence . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

**4 Merging Theorems** **18**

4.1 Conjugate Priors . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18

4.1.1 Convergence of any initial Beta priors on binomial parameter . 20

4.1.2 Convergence of any initial Gamma priors on exponential parameter . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24

4.1.3 Convergence of any initial Dirichlet priors on multinomial parameters . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26

4.2 Savage 1954 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28

4.3 Blackwell & Dubins 1962 . . . . . . . . . . . . . . . . . . . . . . . . 30

4.3.1 Doob's 1953 Martingale Convergence Theorem . . . . . . . . . 30

4.3.2 Merging of opinions with increasing information . . . . . . . . 32

4.4 Schervish & Seidenfeld 1990 . . . . . . . . . . . . . . . . . . . . . . 35

4.5 Uncertain Evidence . . . 36
4.5.1 Huttegger 2015 . . . 37
4.5.2 Nielsen & Stewart 2019 . . . 38
4.6 Formal language and Algorithmic Randomness . . . 39
4.6.1 Gaifman & Snir 1982 . . . 39
4.6.2 Zaffora-Blando 2022 . . . 40

**5 Epistemic Onion** **42**
5.1 Assumptions by Theorem . . . 42
5.1.1 Conjugate Prior-Likelihood . . . 42
5.1.2 Savage . . . 43
5.1.3 Blackwell & Dubins . . . 43
5.1.4 Schervish & Seidenfeld . . . 44
5.1.5 Huttegger . . . 44
5.1.6 Nielsen & Stewart . . . 44
5.2 Comparison . . . 45

**6 Creeping** **47**
6.1 Previous Discussion . . . 47
6.2 Cut and Creep . . . 49
6.3 Application and Discussion . . . 52

**7 Conclusions** **56**

**8 Bibliography** **58**

**9 Appendix** **65**
9.1 R scripts . . . 65
9.1.1 Script: Bayesian Simulations.R . . . 65
9.1.2 Script: Creeping Model.R . . . 73

# List of Figures

1. Figure 1, pg. 19: Extract from Raifa & Schlaifer (1961, pg. 75)
2. Figure 2, pg. 21: Select Priors in the Conjugate setting
3. Figure 3, pg. 23: Learning Development in the Conjugate setting
4. Figure 4, pg. 23: Final Posterior in the Conjugate setting
5. Figure 5, pg. 50: Select Priors for Creeping
6. Figure 6, pg. 50: Classic Learning Development for a True Parameter beyond the support
7. Figure 7, pg. 55: Learning Development for Creeping
8. Figure 8, pg. 55: Final Posterior after Creeping

# List of Tables

1. Table 1, pg. 45: Epistemic Requirements for Merging
2. Table 2, pg. 53: Selected Parameters for Creeping

# 1 Introduction

"*We shall all be Bayesian by 2020*"
(Finetti, 1975, preface)

Unfortunately for Finetti – even four years after 2020 – other theories still prevail and we are not all Bayesian. Why not? How come a consensus between all people on every theory has not emerged yet? This self referential question is actually quite fit to be answered by the Bayesian framework. In philosophy of science, the domain which examines the criteria that lend scientific theories credibility, Bayesianism has established itself as the dominant framework (Bandyopadhyay & Forster, 2011, pg. 5). Bayesian agents are akin to a gold standard for rational learning based on evidence. As Nielsen & Stewart (2018, pg. 2) put it: "Bayesian learning represents a leading contender for an ideal standard of rational revision". Yet, as we will see, rational actors with access to the same evidence do not always reach agreement, challenging *the optimistic thesis about learning* (**TOTAL**) put forward by Nielsen & Stewart (2018, pg. 1):

Rational agents who learn the same evidence resolve disagreements

What are the conditions sufficient for TOTAL to hold? This paper will provide a comprehensive account of the epistemic assumptions behind TOTAL, go through the formal theorems which demonstrate TOTAL, provide detailed proofs for one setting thereof, and finally tackle one assumption seemingly required for TOTAL to obtain with an expanded model. While this paper is primarily situated in Bayesian confirmation theory, the analytic tool set used for the rigour of its conclusions will also borrow from statistics and measure theory. This paper also strives to be accessible. To account for every technical nuance with a corresponding philosophical interpretation is at times beyond the scope of this paper.
The resolution of disagreement, or the consensus between agents considered in this

paper develops in the purely idealistic framework of subjective Bayesianism (Joyce, 2010). The aim is to demonstrate that even towards the heat death of the universe perfectly rational agents may maintain disagreement. The only consideration is the asymptotic development of their disagreement. To this end, any theorems concerned with finite horizons and anything less than complete agreement are disregarded (Kalai & Lehrer, 1994; Lehrer & Smorodinsky, 1996). This long run consensus between ideal agents is called merging of opinions (Blackwell & Dubins, 1962). Any notion of truth here is only a means to the agreement of the agents. As Lindley (2000, pg. 303) puts it: "The apparent objectivity is really a consensus".
The word 'agent' might even be a misnomer for the epistemic agents considered in this paper, as the more so computer-like entities under scrutiny here will not perform any actions. Consequently, giving these entities a capacity to act forms the basis of the motivation behind this field of inquiry. The origin of subjective Bayesianism comes from gambling (Ramsey, 1926). Current Bayesian models back up powerful frameworks in game theory (Kalai & Lehrer, 1993). Most prominently however in the past two decades, Bayesianism is used to explain polarisation and consensus in social settings (Cf. Fitelson, 2009; Nielsen & Stewart 2018; Weatherall & O'Connor, 2020). It is here where Bayesian confirmation theory might most fruitfully help to provide insights into our modern ailments. Terrorism, geopolitical conflict, and environmental disaster are all in part linked to social disagreement. Further, advances in machine learning agents must also prompt a more detailed understanding of social consensus.
To tackle our agenda, the paper is structured as follows. The rest of chapter 1 will summarize the historical development of Bayesian confirmation theory. Chapter 2 will provide some technical preliminaries necessary for a formal understanding of the later theorems. In chapter 3, we construct the epistemic agent by listing the sufficient epistemic requirements for the theorems of the following chapter. Chapter 4 then gets into the thicket of multiple merging of opinions theorems ranging over seven decades. With an understanding of the merging theorems, we can detail in chapter 5 how the development of the field has allowed us to peel off some initial requirements. Chapter 6 presents a novel model to tackle one key requirement. Finally, concluding remarks and an outlook is presented in chapter 7.

## 1.1 Historical Development of Bayesian Confirmation Theory

In order to situate this paper in the academic discussion, a swift history of the field is provided alongside relevant literature. The concepts touched upon in this section will be explained later. Despite what the name Bayesian confirmation theory might suggest, it was mainly the probabilistic methodology which was supplied by its eponym. Confirmation theory can reasonably be traced back to the empirical tradition of Hume. He outlines the clear differences between deductive and inductive justification, and subsequently deplores the lack of justification behind inductive reasoning based on experience (Hume, 1748, pg. 114). Taking up the challenge, confirmation theory is seeking to substantiate how evidence can make hypotheses credible. Rather than defeating Hume's initial problems (Cf. Strevens, 2017, pg. 60), confirmation theory provides a rich discussion on principled inductive reasoning. On the way, the field has extensively intermingled with other fields such as probability theory and economics.

Kvanig (2023) presents a useful delimitation of the development of confirmation theory into three stages: Qualitative, Comparative, and Quantitative. Our concern is the quantitative approach, as it allows for explicit commensurability of the beliefs of different agents, which makes agreement measurable. To this end, probability theory has proven very fruitful. The first initiative to represent the beliefs of agents as subjective probabilities was launched by Ramsey (1926), as a direct critique of the dominant 'objective' probabilities presented by Keynes (1921). The formal framework for probabilities was later provided by Kolmogorov (1936), which directly included Bayes' theorem (pg. 6), hailing all the way back from Bayes (1763). Subjective probabilities and the according personalist view of probability came to fame through their fervent apostles Leonhard Savage and Bruno De Finetti, as espoused in Finetti's "La prévision" from 1937 or Savage's "Foundation of Statistics" (1954). Savage (1954, pg. 3) describes the opposing views to personalism as the necessary or logical view on probability – as seen in Carnap (1950) – and the objectivist view on probability – as seen in Harold Jeffreys (1945; 1961) or Jaynes (1968). From here onwards, confirmation theory has come to be dominated by subjective probabilities and Bayes' theorem. As such, while going through the literature, we have remarked three overlapping periods in Bayesian confirmation theory which may help the understanding of the development. The first period ranges from around the 1920-50s which established the theory in academia. The second period continues from the 1950s-1990s in the anglophone discussion, with many notable successes against rival theories and attacks (Cf. Earman, 1992, pg. 63) as well as merging theorems, but also the spawning of numerous flavours of Bayesianism. Finally, the period up to the present is chiefly concerned with social applications such as agreement and polarisation.

A very broad and immediately accessible introduction to Bayesian confirmation theory is given by Strevens (2017). A more detailed and comprehensive account up to his time is supplied by Earman (1992). A summary of the history and problems of subjective Bayesianism in the academic discussion is run down by Joyce (2010). For an extensive and accessible introduction to Bayesian epistemology, consult the textbook by Titelbaum (2022). Finally, an in-depth overview of all facets in the Bayesian confirmation discussion is collected in Bandyopadhyay & Forster (2011).

# 2 Technical Preliminaries

The Bayesian machinery works only as well as the mathematics which power it. For that reason, a refresher is hereby supplied so as to avoid any confusion in notation. First, the relevant concepts from set- and measure theory are summarized. Measure theory is used exclusively for probabilistic applications here. Any inconsistencies later with the notation here are accidental or due to some very specific context of an application. For further details, comprehensive coverage of measure theory for probability theory is given in Ash & Doléans-Dade (2000). For details on probability theory and applications in stochastic processes, consult Durrett (2019). Regarding Bayesian notation in confirmation theory, this is generally inspired by Strevens (2017) – However, there may also be slight deviations if the use case demands so. A well-versed reader in both probability and confirmation theory can skip this chapter.

## 2.1 Set, Probability, and Measure Theory Preliminaries

Arbitrary capital letters $A = \{a_1, a_2, ..., a_k\}$ will denote **(sub)Sets**; events in probabilistic contexts. The subscript $k$ of the last element denotes the cardinality or **Size** of the set. The **Empty Set** is denoted by $\emptyset$. If $A$ is a **Subset** or equal to $B$, this is denoted as $A \subseteq B$. If $A$ is a **Proper Subset** of $B$ such that at least one element in B is not in $A$, then $A \subset B$. If $A$ is a proper subset of $B$, then the **Complement** of A is denoted by $\bar{A}$. The **Union** of (multiple) disjoint sets $A_1, ..., A_n$ is denoted by $\bigcup\limits_{i=1}^{n} A_i$. The **Intersection** of sets $A_1, ..., A_n$ is denoted by $\bigcap\limits_{i=1}^{n} A_i$.

A (finite) **Partition** $C$ of a set $B$ is a collection of all the non-empty and non-overlapping subsets of $B$; The subsets of the partition are mutually exclusive and collectively exhaustive. For example, for $B = \{a, b, c\}$, a valid partition is $C = \{\{a, b\}, \{c\}\}$ or $C = \{\{a\}, \{b\}, \{c\}\}$. $D$ is a **Refinement** of a partition $C$ if for every subset $d \in D$, there exists a subset $c \in C$ such that $d \subseteq c$. The **Supremum**, or least upper bound of a set denoted by *sup*, is the least value that is still greater or equal to every other element in the set. The **Infimum** which is the inverse concept thereof will be denoted by $inf$.

A **Probability Space** consists of $(\Omega, \mathfrak{F}, \mu)$. $\Omega$ is the **Sample Space**, i.e the set of all possible atomic outcomes which may obtain. The **Sigma Algebra** $\mathfrak{F}$ is a collection of subsets of $\Omega$ which is closed under countable unions and complementation, including the empty and sample space. These special sets are denoted in the Fraktur font. The subsets of $\mathfrak{F}$ can be called 'events'. Some event containing only one outcome is called an atomic or elementary event. When $\Omega$ is countable, $\mathfrak{F}$ is usually the power set $\mathfrak{P}$. For example, if $\Omega = \{a, b, c\}$, then $\mathfrak{F} = \sigma(\Omega) = \mathfrak{P} = \{\{\emptyset\}, \{\Omega\}, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}\}$. A sub-sigma algebra $\mathfrak{B}$ is a subset of $\mathfrak{F}$, but also fulfills all conditions of a sigma algebra on its own. A **Filtration** is a sequence of sigma algebras such that for all $k \leq n$, $\mathfrak{F}_k \subseteq \mathfrak{F}_n$. A **Random Variable** $X$ is a sigma algebra measurable function such that $X : \Omega \to \mathbb{R}$, whereby the outputs are often called (quantifiable) realisations of the random variable, denoted by $x$. The **Probability Density Function** (PDF) $f(x)$ assigns densities to ranges (or probabilities to elements in the discrete case: probability mass function) to the output $x$ of the random variable, such that if the domain of $x$ ranges across the support $[a, b] \subseteq \mathbb{R}$, then necessarily $\int_a^b f dx = 1$, and $f$ is strictly non-negative. A PDF is not to be confused with a **Probability Measure** $c : \mathfrak{F} \to [0, 1]$ such that $c(\Omega) = 1$. While it is unusual to have $c$ denote a measure, this letter will be used here to emphasise that may be some agent's credence, more on this later. A (probability) measure $c$ is absolutely continuous to another measure $\mu$ if for all events $A \in \mathfrak{F}$, $c(A) > 0 \implies \mu(A) > 0$, or succinctly: $c \ll \mu$. All measures are **Countably Additive** unless explicitly not declared to be so. This means for a countable set of disjoint events $A_i$, it holds that $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$, which also implies the simpler notion of **Finite Additivity**: $\mu(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i)$. If for some sort of theorem given a measure $\mu$ , e.g a convergence theorem like the strong law of large numbers (Cf. Ash & Doléans-Dade, 2000, pg. 235-), this theorem fails to hold only for subsets of the sample space $A \subset \Omega$ where $\mu(A) = 0$, then this theorem holds **Almost Everywhere** (a.e). If some event is assigned probability 1, then this event obtains **Almost Surely** (a.s). As such, the two concepts are analogues for measure and probability theory respectively, such that Doob (1971, pg. 456) writes: "One of the most noticeable distinctions between probability and measure theory is that a probabilist frequently writes 'almost surely' where a measure theorist writes 'almost everywhere' ". Generally, the two terms mean that theorems of interest hold in those areas deemed possible.

The final two concepts are relevant measure theoretic technicalities, but their details are beyond the scope of this paper. One technical assumption for this paper is that all probability spaces allow for **Regular Conditional Probabilities**. This means that for some measure $\mu$, if $\mathfrak{B}$ is a sub-sigma algebra of $\mathfrak{F}$, then there exist countably additive conditional measures $\mu(\cdot|\mathfrak{B})$ over $\mathfrak{F}$ for every $\omega \in \Omega$. Finally, the **Radon-Nikodym Theorem** guarantees that given a measurable space $(\Omega, \mathfrak{F})$, and two measures $q \ll \mu$, then there exists a $\mathfrak{F}$-measurable function $g$ such that for any subset or event $A \in \mathfrak{F}$: $q(A) = \int\limits_A g \; d\mu$.

## 2.2 The Bayesian Machinery

This section will detail the Bayesian machinery which is employed in this paper alongside the notation of its components. There is a running gag in Bayesian literature that there are more forms of Bayesianism than actual Bayesians (Earman, 1992, pg. 33). I.J Good supposedly calculated 46'656 types of Bayesianism. Given that the literature on Bayesian confirmation theory spans nearly a century at this point, we will try to formalize the notation as it is employed in this paper. Any deviations from the notation here are related to the exact context of the applications. This section will not dive into the epistemic aspects of each component.

To start, we need (static) subjective probabilities. We consider a countable set of **Hypotheses**, which can be thought of as mutually exclusive propositions, such that $h_i \in H$. An epistemic agent has a **Credence** $C$ ranging over these hypotheses, which can be modeled as a probability mass function. This means the credence behaves according to the probability axioms (Cf. Kolmogorov, 1933). Therefore,

$C(h_i) \geq 0$, $C(h_i \cup h_k) = C(h_i) + C(h_k)\ \forall i \neq k$, and $C(H) = 1$.
Now the epistemic agent iteratively receives **Evidence** data $e_1, e_2, ..., e_n$ whereby the subscript is often dropped. In statistical contexts, these could be the realizations of a random variable. Accordingly, we are interested how much the evidence confirms the hypotheses. To that end, the agent becomes dynamically coherent by making use of Bayes' theorem (Cf. Bayes, 1763, pg. 378, 381). The probability of interest is the conditional probability of the hypotheses given the evidence:

$$C(h_i|e_j) = \frac{C(h_i \cap e_j)}{C(e_j)} = \frac{C(e_j|h_i)}{\sum_{i=1}^{n} C(e_j|h_i) \cdot C(h_i)} \cdot C(h_i)$$

Whereby the latter formulation is the widely known form of Bayes' Theorem. The denominator is simply rewritten according to the law of total probability. The mechanics of this classic Bayesian inference rely on three components:
**The prior:** $\mathbf{C(h_i)}$ which is the unconditionalized subjective probability ranging over a set of hypotheses $h_i$. It is possible to add a subscript to the credence function according to the index of the evidence upon which it has been conditionalized. As such, the first prior of some iterative inference process can be written as $C_0$, which could be considered the 'tabula rasa' prior in some settings.
**The likelihood function:** $\mathbf{C(e_j|h_i)}$ which gives the likelihood of $e_j$ given some hypothesis. If hereby some objective likelihood is considered, then this component can be rewritten as $\mathbf{P(e_j|h_i)}$.
**The posterior:** $\mathbf{C(h_i|e_j)}$ which is the normalised product of the prior and the likelihood function. The posterior bears all the same probabilistic functionalities as the prior, and is used as the prior of the next iteration.
A generalized version of classic conditionalizing comes in the form of **Jeffrey Conditionalization** (Jeffrey, 1965, pg. 164). This form is useful when evidence is uncertain, which will appear in later merging results. Suppose you have a finite set of different evidence data $e_1, e_2, ..., e_n$. The agent is uncertain which one of these mutually exclusive data has obtained this iteration, but knows that one has obtained almost surely. The posterior can then be written as a weighted average:

$$C_1(h_i) = \sum_{j=1}^{n} C_0(h_i|e_j) \cdot C_0(e_j)$$

Which turns back into classic conditionalization if there is certainty as to which datum obtained. Further on the classic conditionalization, in a continuous setting where $h \subset H$, the conditionalization could be rewritten as:

$$C(h_i|e_j) = \frac{P(e_j|h)}{\int_H P(e_j|h)C(h)\ dh} \cdot C(h)$$

Finally, making some assumptions about the likelihood and the evidence, we can succinctly rewrite the entire conditionalization over multiple iterations as a proportionality as there are no unique components in the normalizing denominator.

$$C(h_i|e_j) \propto \prod_{j=1}^{n} P(e_j|h_i) \cdot C_0(h_i)$$

While TOTAL offers a qualitative framework for asymptotic agreement, **Agreement** between agents can be formalized in different ways. In simple models, the posterior of different priors might display **Convergence** for one defined value almost surely: $C_\infty(h_t) = 1$. For comparisons between two measures $c$ and $\mu$ conditionalized on $n$ iterations of evidence, **Merging** means the total variation distance converges almost surely to zero:

$$\lim_{n\to\infty} d(c,\mu) = \lim_{n\to\infty} \sup_{h_i \in H} |c_n(h_i) - \mu_n(h_i)| = 0$$

# 3 An Epistemic Agent

To begin, we will construct an epistemic agent equipped with an analytic arsenal of assumptions. This chapter will detail these preliminary assumptions required by our agents before we can delve into their asymptotic agreement. These assumptions are akin to an onion with layers; Later in chapter 5, the assumptions are progressively peeled away as subsequent theorems demand less of their agents. As such, the contents of this chapter will culminate in chapter 5.
First, note that the Bayesian framework is a tightly knit construction of entangled assumptions, which makes it difficult to analytically divorce the components of the framework and go through them in an ordered manner. Nevertheless, an attempt will be made. Further, this chapter will not argue for or against any of these requirements. The assumptions are simply required for a critical reading of the merging theorems in chapter 4. Wherever resistible, the roads not taken are beyond the scope of this paper in order to maintain focus.
As our goal is to elucidate the conditions for TOTAL, we first discuss why the basic assumptions of the agent are rational. The core of our epistemic onion is formed by formal language, which might not the be the only way of approaching these epistemic requirements; From a philosophically anthropocentric view this approach is convenient, but from a statistical science view we might start mathematically straight away. Then, after imbuing our agents with an account of rationality, we move to understanding agreement between agents in the context of subjective probabilities. Finally, the agents are exposed to iterative evidence which they learn from.
An important note at this point is that this onion model of a Bayesian agent which will be constructed here prescribes merely sufficient assumptions, but perhaps necessary requirements have been missed. Further, the method and order of what assumptions are elucidated is heavily influenced by the tradition of analytic philosophy.

## 3.1 Bayesian Agents as Propositional, Logical, Rational

We start at first with an epistemic agent as basic as conceivably possible. Consider the **stream of experience** this agent undergoes. The very first thing it must do is to somehow **partition** this stream, i.e the stream must be separable by the agent. The nature of this partition is left quite open for our context of personalism. The elements of this partition are objects. These objects are what **propositions** – in a logical context – or events – in a probabilistic context – deal with (Finetti, 1975, pg. 5). First, regarding the logical context, these objects bear a **binary truth value** (pg. 25). This enables the usage of logical operations such as connectives etc. Further, they are events when the agent considers the potential 'occurrence' or 'obtainment' (Savage, 1954, pg. 10) of these propositions. Further dimensions beyond truth are not considered. The expression of these propositions is through sentences of a specific **formal language** (Earman, 1992, pg. 35). The agent is at all times aware of all logical truths entailed in the language. We now have enabled the agent to play with **propositional first order logic**. The (true) world may be considered the set of all (true) propositions.
The next requirement is that this agent has a general **propositional attitude** towards the considerable propositions, which should aim to represent what the world is descriptively like, whatever this may mean to the agent. These relevant propositional attitudes here are called **doxastic attitudes**. Doxastic attitudes include belief, certainty, and credence (Titelbaum, 2022, pg. 3). What does it mean for these doxastic attitudes, for example the beliefs of an agent, to be '**rational**'? Titelbaum (2022, p. 9) proposes necessary conditions such that i) the beliefs are logically consistent with each other, and ii) the beliefs form a logical closure. This means that for the agent to be rational, its beliefs need to be coherent in some sense. The marriage of rationality and logical consistency is so fundamental to Bayesian inference that Savage (1954, pg. 7) simply states his account of rationality as follows: " 'rational' means logical, there is no live question". During the genesis of personal probability, Ramsey (1926, pg. 191) also equates rationality with logic, further stating: "We found that the most generally accepted parts of logic, namely, formal logic, mathematics [...] are all concerned simply to ensure that our beliefs

are not self-contradictory". This approach to rationality will be sufficient for the purposes of our agent.
A minor tangent might elucidate the previous elaboration. Approaches to rationality could be made in much more general strides (Cf. Kvanig, 2014, pg. 48). Kvanig takes a value-driven approach to determining what rationality is and how we should think about rationality generally. Inspired by James (1897), Kvanig (2014, pg. 10) notes: "Human beings are motivated by two primary concerns, a concern for not being duped and a concern for not missing out on something important". Such an approach to rationality does not explicitly necessitate the use of a Boolean algebra as employed by Ramsey, Finetti, or Savage. Other 'exotic' approaches might provide exciting results. For example, Bandyopadhyay & Forster (2011, pg. 44) explain how different Indian paradigms of logical thought exist couched within ancient epics as old as India herself. The esoteric cloaking of those mathematical findings might lead Western observers to deem Indian philosophy as "mystical and irrational". As Raju (2011, pg. 17) notes, there are have been attempts to construct logically consistent systems of probability with multi-valued logic. However, the Boolean first order logic shall suffice for now.
What we then require is that the agent can make a transitive **belief-ranking** according to some confidence metric. This gives us the two core tenets of Bayesian epistemology (Titelbaum, 2022, pg. 12):

> 1. Agents have doxastic attitudes that can usefully be represented by assigning real numbers to claims.
>
> 2. Rational requirements on those doxastic attitudes can be represented by mathematical constraints on the real-number assignments closely related to the probability calculus.

A credence then fulfills these requirements by assigning a **subjective probability** to the considered propositions, or now called hypotheses. To this end, the agent's language also needs to accommodate at least a basic first order language for arithmetic (Cf. Gaifman & Snir, 1982, pg. 500). In order for credences to be rational, they must obey the axioms of probability as proclaimed by Kolmogorov (1933, pg. 2). To effectively run the Bayesian machinery with subjective probabilities, we also assume that the agent is "logically omniscient" (Earman, 1992, pg. 121), meaning that the agent is always aware of all logical truths of the language upon which the credence is defined. Up until this point most Bayesians have agreed, but from here on out they may scatter into all their different flavours (Bandyopadhyay & Forster, 2011, pg. 5).

An additional assumption will require the hypotheses to be strictly **observational hypotheses** and not theoretical hypotheses, in order to circumnavigate a large dispute on the underdetermination of theory by observational evidence (Earman, 1992, pg. 149). Basically, there are epistemic and logical difficulties in exactly determining how observational outcomes directly relate to a larger theoretic framework. A form in which this dispute is being waged is by tackling an analogue of the Quine-Duhem thesis for Bayesian confirmation theory (Strevens, 2017, pg. 108). A workaround is feasible here, but to retain sufficient requirements we will stick to observational hypotheses only.

Finally, to really gain mathematical clarity for the results, we require the credence to be carried over into a **probability space triplet** of measure theory $(\Omega, \mathfrak{F}, \mu)$. Much of this work has cleanly been done by the Gaifman & Snir (1982), which will be covered in more detail later. To already get a taste, the sample space is the set of all models of a language, the sigma algebra is generated by sets of models, and the countably-additive measure is defined over sentences (Earman, 1992, pg. 37). While **countable additivity** might prove some difficulties for both formal and ordinary language (Cf. Hawthorne, 2011, pg. 341), its mathematical applications make it highly useful for the merging results. For example, the Strong Law of Large Numbers relies on countably additivity (Cf. Earman, 1992, pg. 62).

We now have a rational agent using a formal language who assigns subjective probabilities over its hypotheses at a given moment. As such, this agent is now statically coherent.

## 3.2 Immediate agreement

We can now look into two agents with the previous assumptions. When considering agreement or consensus, we imagine that there is no discussion or exchange between the two agents; the credences they assign are simply readable from some omniscient perspective, as if we could read a brain scan which directly displays the beliefs of an agent. That there is a precise **real number assignment** to credences allows for commensurability, i.e the relation between the credences of agents becomes explicit in a very Kelvin sense. In a more broad sense, the agents don't even need to assign rational numbers for each hypotheses, but the quantification needs to simply be empirically adequate to explain their behaviour. Titelbaum (2022, pg. 14) quips in response to opinions unfavourable of this quantification that this attitude is like "refusing to measure gas samples with numerical temperature values because molecules don't fly around with numbers pinned to their backs". For our purposes, we will assume for sufficiency that the agents do have numbers in their heads. For the formal measure theoretic context which our agents comply with, (dis)agreement is measured through the **total variation distance**.

Without considering any common evidence yet, can two such agents come to a consensus on all propositions? Do they agree? If they are rational, as claimed before, can they even disagree? Savage (1954, pg. 67) famously stated:

> "The personalistic view incorporates all the universal acceptable criteria for reasonableness in judgement known to me [...] The criteria incorporated in the personalistic view do not guarantee agreement on all questions among all honest and freely communicating people, even in principle. That incompleteness [...] does not distress me, for I think that at least some of the disagreement we see around us is due neither to dishonesty, to errors in reasoning, nor to friction in communication."

Modern positions such as those by Kvanig (2014, pg. 98) or Kelp (2012) also propose that rational disagreement is possible. Even if agents communicate – which they do not in this paper – they may still rationally disagree (Cf. Conee, 2010, pg. 69–). This philosophical discussion is a huge can of worms which we will not enter here, and in line with subjective probabilism/personalism, the account of rationality used here does allow for disagreement between agents. An analogue of the opponents to rational disagreement come to Bayesian confirmation theory in the form of objective Bayesians, which want to constrain the initial or prior probabilities over hypotheses (Jeffreys, 1945; Jaynes, 1968). Their mantra can be summarized by Jaynes (1968, pg.3): "In two problems where we have the same information, we should assign the

same prior probabilities". The application of some Leibnizian principle of insufficient reason to get some 'ignorant', 'unassuming', or 'invariant' prior has proven extremely tricky however, and an objective prior remains highly disputed (Earman, 1992, pg. 57; Joyce, 2010, pg. 9; Norton, 2011, pg. 408). The issues thereof relate to the *Bertrand paradox*. To get a quick taste of the discussion, consider a machine which produces cubes of unknown dimensions. Assume some uniform distribution over a reasonable length one side could take. This should be logically equivalent to making an assumption about the volume as we are talking about a cube. So the prior over some real range $[a, b]$ of lengths $L$ is $C(l) = \frac{1}{b-a}$, but over the volume $V$ in the range $[a^3, b^3]$ the equivalent prior is $C(v) = \frac{v^{-2/3}}{3(b-a)}$, which is clearly considers smaller volumes more likely. To circumvent this entire debate again, we will not make any assumptions to constrain the prior; this is a strength of subjective Bayesianism.
If there are no restrictions on the prior – besides obeying probability calculus of course – then the agents will most likely disagree on a lot of propositions initially. Further, the agents might not even agree on what events could possibly obtain; their credences might not be mutually absolutely continuous. Many different terms are used for this condition: "radical disagreement" (Nielsen & Stewart, 2018, pg. 13), "discrepant" (1990, Schervish & Seidenfeld, pg. 334), or "dogmatic" (Earman, 1992; Gaifman & Snir, 1982; Zaffora-Blando, 2022). We have been and are going to continue to use 'dogmatic'. We further require the agents to be **equally dogmatic**. Given then a subjective prior, it is still apparent that hopes for initial agreement is in vain.

## 3.3 Iterative evidence

Could statically coherent beliefs ensure agreement between two agents? Ramsey (1926, pg. 27) replies: "But obviously not, this is not enough; we want our beliefs to be consistent not merely with one another but also with the facts". Perhaps now, the agents can reach a consensus, not through discussion with one another but simply by observing the same **evidence** – data which appears as the same propositions with the same truth value to both agents. As such, we are now entering dynamic coherence by conditionalizing credences on evidence (Cf. Levi, 1980, pg.

104-; Skyrms, 1987). The nature of this evidence comes as events, or in clear-cut sentences for propositional logic. The obtainment of these events is assigned a clear positive truth value (Finetti, 1975, pg. 25), which must be the same value for both agents. For the simple case, this means only one event or proposition obtained, and it is the same event for both agents. This of course requires, as also evident in the following conditional probability definition, that the event needs to have been granted a positive prior credence by both agents. How exactly the observational process works here is a black box for now – we simply assume that propositions are deterministically feed to the agents like a tape into some Turing machine. To generalize, this is an account of propositionalism about evidence (Kvanig, 2014, pg. 57).

The Bayesian **conditionalization** of the credences occurs then as defined (Earman, 1992, pg. 36):

$$C(h|e) \equiv \frac{C(h \cap e)}{C(e)} = \frac{C(e|h)}{C(e)} \cdot C(h), \text{ if } \; C(e) \neq 0$$

In its most generalized form, this can be formulated as Jeffrey conditionalization. The clean cut form of evidence as a proposition also implies that the conditionalization is iterative. For sufficiency, the order of this iteration is fixed and the same for both agents, as there are certain problems that can occur in specific contexts if exchangeability is permitted (Joyce, 2010, pg. 45). We now assume that an infinite number of these ordered **discrete iterations** occur.

By iteratively incorporating evidence into its credence, the agent maintains memory. After many iterations, certain hypotheses are going to inductively cement themselves as more confirmed than others. As Hume (1748, pg. 132) already formulated in qualitative terms:

> “ When we transfer the past to the future [...] as a great number of views [observations] do here concur in one event [hypothesis], they fortify and confirm it to the imagination, beget that sentiment which we call belief, and give its object the preference above the contrary event, which is not supported by an equal number of experiments, and recurs not so frequently to the thought in transferring the past to the future.”

Hume emphasizes the necessity of assuming the **uniformity of nature** for induction, which underpins Bayesian confirmation (Cf. Strevens, 2017, pg. 62; Zaffora-Blando, 2022), ensuring agents' temporal cohesion for reliable confirmation. In a similar vein, it is also assumed that the act of conditionalizing does not impact the way the evidence is produced. The propositional knowledge generated or evidence perceived is completely independent of our accumulated learning. This is called the **martingale** assumption, the technical form and relevance of which will be elaborated upon in Chapter 4.

Moving on to the likelihood, how does an agent set $C(e|h)$? There is a large discussion on how this probability is to be set, and pure personalists consider no rational constraints on this term. As Strevens (2017, pg. 34) notes, we are constraining this probability according to the **probability coordination principle** (Cf. Joyce, 2010, pg. 55). This assumption leads to $C(e|h) = P(e|h)$, where the latter is the 'objective' or physical likelihood of $e$ given $h$. For now, this principle ensures that both agents select the same 'right' likelihood function.

For our purposes, the key is not asymptotic *convergence* to certainty but *merging* – the variational distance between measures reaching zero (Blackwell & Dubins, 1962, pg. 883) – as agents update their hypotheses. We don't need the convergence of credence to a specific value – it is conceivable that the credences of the two agents oscillate with each iteration, but there is still consensus (Cf. Earman, 1992, 245). However, convergence to certainty famously yields a first pass on merging of opinions. If agents are *certain* of the same hypotheses, then they also consequently agree, but not vice versa.

Finally, all merging occurs only almost surely, meaning with $C$-probability 1, which should not be equated with the merging concretely occurring. As Kolmogorov (1933, pg. 5) already noted conversely, an event assigned $C$-probability 0 might still obtain. This is akin to the anecdote about infinite monkeys on typewriters –. one will almost surely write Shakespeare's Hamlet, but good luck finding that monkey. Keynes (1921), a constrained objective Bayesian, may again retort that we're all dead in the long run. As Earman (1992, pg. 148) famously puts it: "'Almost surely' sometimes serves as a rug under which some unpleasant facts are swept".

With the above considerations, we delve into multiple merging theorems developed over decades. Chapter 5 will examine the extent to which the constructed epistemic onion can be peeled.

# 4 Merging Theorems

With the earlier assumptions in mind, we can now wield the concepts from the technical preliminaries of chapter 2 to formulate multiple merging theorems. These theorems start from from purely statistical models and develop into more humanized models. This chapter will lend the technical models and proofs the main stage. The epistemic assumptions are reexamined in chapter 5. All simulations are done on R and the corresponding scripts are in the appendix under section 'Bayesian Simulations.R'

## 4.1 Conjugate Priors

A prior and a likelihood function of the same family of distributions allow for closed-form solutions of the posterior (Raifa & Schlaifer, 1961). In such a case, the prior is *conjugate* to the likelihood. This makes asymptotic analysis incredibly easy, but also restricts the analysis to a set of compatible functions. As such the conjugate priors setting is a specific cookie-cutter application of general Bayesian confirmation theory.

In this setting, the notation is slightly adapted to better suit the relationship between the likelihood function and the prior. The evidence is assumed to be observed realisations of a random variable distributed according to some true parameter '$\theta_t$' (Cf. Raifa & Schlaifer, 1961, pg. 70). Let $n$ denote the number of observed evidence data. The set of hypotheses $H$ are now denoted by $\Theta$, because the hypotheses are essentially mutually exclusive propositions which state that the true, objective parameter is $\theta$; These are called statistical hypotheses. It is important to reiterate however, that the general purpose of this paper is not to demonstrate convergence of beliefs to some concept of 'truth', but that this is merely a conduit for the merging of the subjective priors of different agents.

The relations between the following selected conjugate prior-likelihood pairs are well known, as seen in Figure 1 below. What follows are our complete proofs showing that the posterior will obtain a constant expectation and zero variance as $n \to \infty$. In other words, the posterior $C(\theta|e)$ will collapse onto a single point regardless of what parameters determined the initial shape of the prior $C(\theta)$. As such, the Bayesian estimation procedure is consistent.

**Proposition:**

$$\lim_{n\to\infty} E[C|e] = \theta_t \quad \text{and} \quad \lim_{n\to\infty} Var[C|e] = 0 \ \ \forall\, C(\theta)$$

**Figure 1.** Extract from Raifa & Schlaifer (1961, pg. 75):

**Table 3.1**

**Distributions of Statistics and Posterior Parameters**

| Process | Prior Distribution | Experiment | Distribution of Statistic: Conditional | Distribution of Statistic: Marginal | Distribution of Posterior Parameter |
|---|---|---|---|---|---|
| Bernoulli | beta | $\tilde{r}\|n$ | binomial (9.2.2; 7.1) | beta-binomial (9.2.3; 7.11) | — |
| | | $\tilde{n}\|r$ | Pascal (9.3.2; 7.2) | beta-Pascal (9.3.3; 7.11) | — |
| Poisson | gamma-1 | $\tilde{r}\|t$ | Poisson (10.3.2; 7.5) | negative-binomial (10.3.3; 7.10) | — |
| | | $\tilde{t}\|r$ | gamma-1 (10.2.2; 7.6.2) | inverted-beta-2 (10.2.3; 7.4.2) | — |
| Normal | $h$ known; $\tilde{\mu}$ Normal | $\tilde{m}\|n$ | Normal (11.4.1; 7.8.2) | Normal (11.4.2; 7.8.2) | Normal (11.4.3; 7.8.2) |
| | $\mu$ known; $\tilde{h}$ gamma-2 | $\tilde{w}\|\nu$ | gamma-2 (11.2.1; 7.6.4) | inverted-beta-2 (11.2.2; 7.4.2) | inverted-beta-1 (11.2.3; 7.4.1) |
| | Normal-gamma | $\tilde{m}, \tilde{v}\|n, \nu$ | *See* 11.6.1 | 11.6.2, 11.6.3 | 11.7.1, 11.7.2 |
| Multinormal | $h$ known; $\tilde{\boldsymbol{\mu}}$ Normal | $\tilde{\mathbf{m}}\|\mathbf{n}$ | Normal (12.2.1; 8.2) | Normal (12.2.2; 8.2) | Normal (12.3.1; 8.2) |
| | Normal-gamma | $\tilde{\mathbf{m}}, \tilde{v}\|\mathbf{n}, \nu$ | *See* 12.5.1 | 12.5.2, 12.5.3 | 12.6.1, 12.6.2 |
| Regression | $h$ known, $\tilde{\boldsymbol{\beta}}$ Normal | $\tilde{\mathbf{b}}\|\mathbf{n}$ | Normal (13.3.1; 8.2) | Normal (13.3.2; 8.2) | Normal (13.4.1; 8.2) |
| | Normal-gamma | $\tilde{\mathbf{b}}, \tilde{v}\|\mathbf{n}, \nu$ | *See* 13.6.2 | 13.6.3 | 13.7.1, 13.7.2 |

### 4.1.1 Convergence of any initial Beta priors on binomial parameter

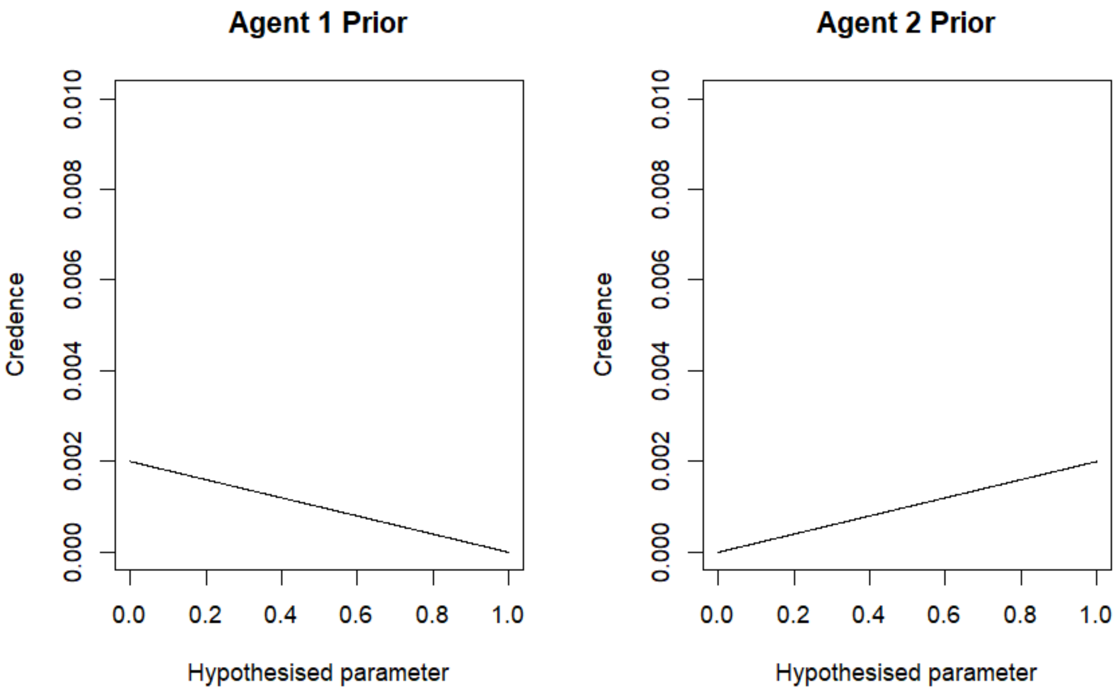
**Example:** While strolling through the garden, Adam and Eve stumble upon a coin. One side is head, one is tails, but they are unsure if the coin is fair or weighted. In their edenic naivety, they assume that upon flipping the coin, one side might be favoured over the other. They reflect decide how likely they perceive each potential weighting a priori. Then, they start flipping the coin and observing the outcomes, each time noting down the result to then form a sequence of heads and tails. The coin does not wear out in this process. For this endeavour they have endless time. Will they eventually completely agree on the weighting of the coin?

**Model:** The prior belief is given by the distribution $Beta(a, b)$ s.t $a, b > 0$. Accordingly, $C(\theta) = f(\theta; a, b)$ The support of this PDF is the unit interval. Parameters $a$ and $b$ determine the shape of the PDF. The evidence $e$ is a sequence of independently and identically distributed (i.i.d) Booleans. In this sequence, $k$ will denote the number of successes or ones and $n$ shall be the length of the sequence. Consequently, an agent considers different infinite sequences of binary outcomes $\omega \in \Omega$ and the associated fraction of 1's: $\theta \in \Theta$, or formally the Cartesian product $\Omega \times \Theta$ (Cf. Sterkenburg & De Heide, 2021, pg. 3). The likelihood function is the binomial probability mass function $Pr(k, n, p = \theta)$. As such, the unit interval support of the Beta distribution is a very intuitive representation of all potential parameters $\theta$ of the Binomial distribution.

The PDF of the Beta distribution is given here by: $f(\theta; a, b) = c \ \cdot \ \theta^{a-1}(1-\theta)^{b-1}$ where $c = \frac{(a+b-1)!}{(a-1)!(b-1)!}$

For two agents with prior parameters $a = 1, b = 2$ and $a = 2, b = 1$, the following prior seen on Figure 2 is generated.

**Figure 2:**




**Proof of proposition:**

$$C(\theta|e) = \frac{\binom{n}{k}\theta^k(1-\theta)^{n-k}}{\int_0^1 c\cdot\theta^{a-1}(1-\theta)^{b-1}\cdot\binom{n}{k}\theta^k(1-\theta)^{n-k}\,d\theta}\cdot C(\theta)$$

$$= \frac{\cancel{c\cdot\binom{n}{k}}\theta^{k+a-1}(1-\theta)^{n-k+b-1}}{\cancel{c\cdot\binom{n}{k}}\int_0^1\theta^{a+k-1}(1-\theta)^{n-k+b-1}\,d\theta}$$

As seen in Abramowitz & Stegun (1972) on the Beta function, the denominator is an Euler integral of the first kind, allowing the denominator to be rewritten as a Beta function.

$$= \frac{\theta^{k+a-1}(1-\theta)^{n-k+b-1}}{B(a+k, b-k+n)}$$

Further, the Beta function can be rewritten in terms of a ratio of Gamma functions. The Gamma function basically allows factorials to be extended to non-integers, and for a given integer $\alpha$ can be written as $\Gamma(\alpha) = (\alpha - 1)!$

$$= \frac{\theta^{k+a-1}(1-\theta)^{n-k+b-1}}{\frac{\Gamma(a+k)\cdot\Gamma(n-k+b)}{\Gamma(a+n+b)}}$$

$$= \frac{\theta^{k+a-1}(1-\theta)^{n-k+b-1} \cdot (a+n+b-1)!}{(a+k-1)!(n-k+b-1)!}$$

$$= \frac{(a+n+b-1)!}{(a+k-1)!(n-k+b-1)!} \cdot \theta^{k+a-1}(1-\theta)^{n-k+b-1}$$

$$= f(\theta; a+k, b+n-k) = C(\theta|e)$$

As such, we have a closed form solution for the posterior. Now, it remains to be seen if the expectation of the posterior asymptotically collapses onto a single point with a variance of 0. The expectation and variance of the Beta distribution are well known.

$$\lim_{n\to\infty} E[C|e] = \lim_{n\to\infty} \frac{a+k}{a+k+b+n-k} = \underline{\underline{\frac{k}{n}}}$$

$$\lim_{n\to\infty} Var[C|e] = \lim_{n\to\infty} \frac{(a+k)(b+n-k)}{(a+k+b+n-k)^2(a+k+b+n-k+1)}$$

$$= \lim_{n\to\infty} \frac{(a+k)(b+n-k)}{(a+b+n)} \cdot \frac{1}{a+b+1+\mathbf{n}} = \underline{\underline{0}}$$

Thus it is proven that the asymptotic result of the posterior is independent of the initial parameters $a$ and $b$ and only dependent on the data. Simulating this process with the earlier shown priors nicely reflects the theoretical results. In this simulation, a true parameter of 0.3 is set, and 800 iterations are run. On Figure 3, each iteration is differently colored to display the learning development.

As displayed on Figure 4, the final posterior narrows down near the true parameter. The hypothesis with highest credence is indexed by the red horizontal line. For agent 1 and 2 this line is at 0.308 and 0.31 respectively.

**Figure 3:**

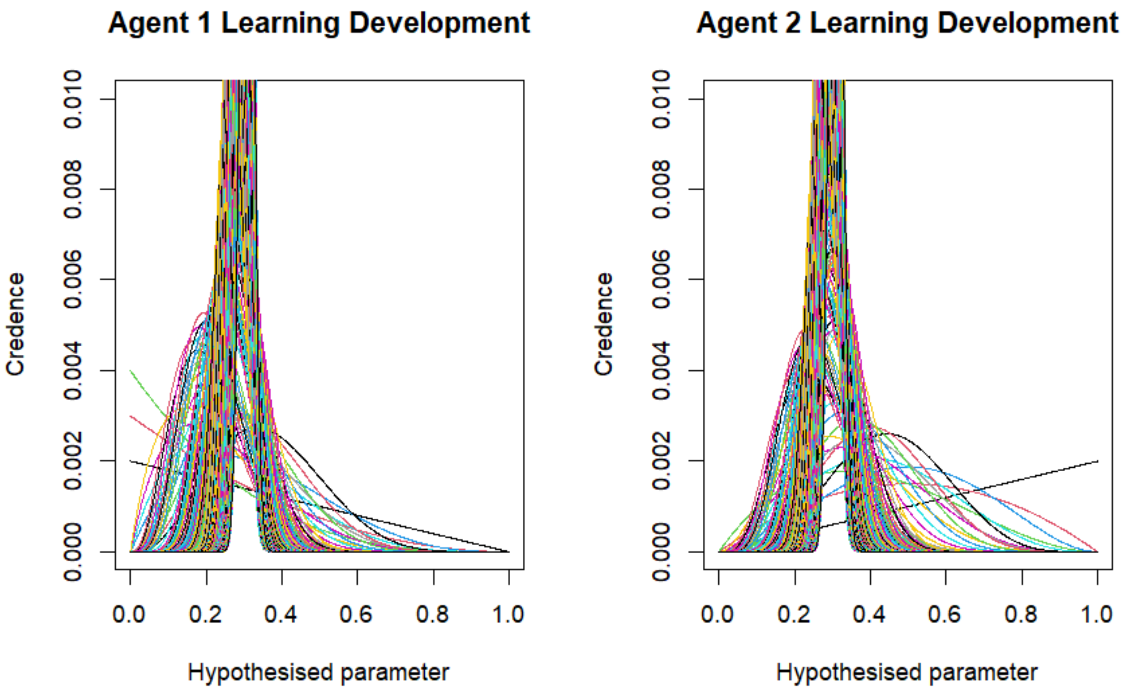
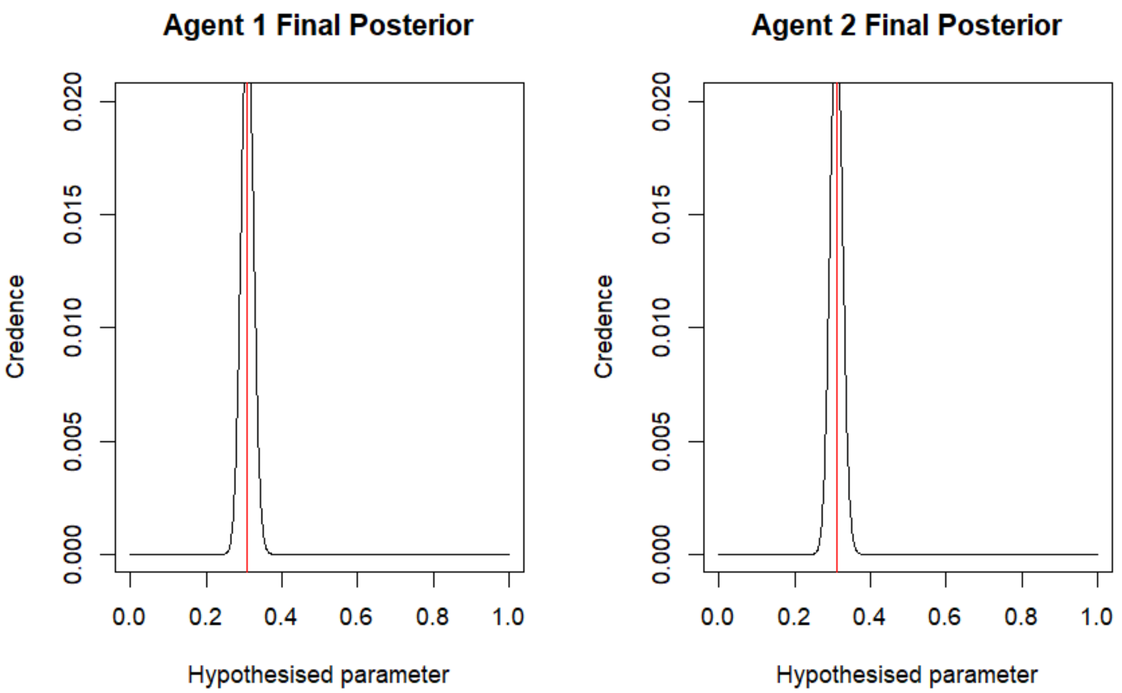



**Figure 4:**



### 4.1.2 Convergence of any initial Gamma priors on exponential parameter

**Example:** Upon perfectly agreeing on the weighting of the coin, Adam and Eve continue to a tree, where many fruit flies have begun to feast on dropped fruit. They wonder how long one of these flies lives, and initially disagree heavily. Having infinite time at their disposal, they go about observing and noting down the exact lifetimes of many different fruit flies all across the garden. Will they in the end agree on what the life expectancy of a fly is?

**Model:** The prior belief is given by the distribution $Gamma(a, b)$ s.t $a, b > 0$. Therefore, $C(\theta) = f(\theta; a, b)$. Parameters $a$ and $b$ again determine the shape. The support of the probability density function ranges across all positive real numbers including zero. The evidence $E = \{e_1, e_2, e_3, ..., e_n\}$ is a set of i.i.d real positive numbers of size $n$. The likelihood function is the exponential density function: $\theta e^{-\theta \cdot e_i}$. To avoid any unfortunate confusion in notation,it is important to clarify that the base $e$ denotes Euler's number and the $e_i$ in the exponent denotes an observed value. The PDF of the Gamma distribution is given here by: $f(\theta; a, b) = \frac{\theta^{a-1} e^{-b\theta} b^a}{\Gamma(a)}$, where the denominator is the Gamma function.

**Proof of proposition:**

As we are simply interested in an asymptotic result, we can calculate the posterior by ignoring the normalising denominator while conditionalising. The denominator only ensures that the posterior is a valid PDF, but doesn't affect the general shape of the posterior. As only the asymptotic point convergence is of interest and not any specific probabilities on the way, we can eschew the denominator and further normalising constants in the proof.

$$C(\theta|e_i) \propto \theta e^{-\theta \cdot e_i} \cdot Gamma(\theta; a, b)$$

$$C(\theta|E) \propto \prod_{i=1}^{n} \theta e^{-\theta \cdot e_i} \cdot Gamma(\theta; a, b) \ \propto \ \theta^n e^{-\theta \sum_{i=1}^{n} e_i} \cdot \theta^{a-1} e^{-b\theta}$$

$$\propto \ \theta^{n+a-1} e^{-\theta(\sum_{i=1}^{n} e_i + b)}$$

$$\rightarrow C(\theta|E) = Gamma(\theta; a+n, \sum_{i=1}^{n} e_i + b)$$

Now, to check wether this closed form posterior collapses onto a single point with zero variance, independently of the prior parameters $a$ and $b$

$$\lim_{n\to\infty} E[Gamma(\theta; a+n, \sum_{i=1}^{n} e_i + b)] = \lim_{n\to\infty} \ \frac{a+n}{\sum_{i=1}^{n} e_i + b}$$

$$= \lim_{n\to\infty} \ \frac{a}{\sum_{i=1}^{n} e_i + b} + \frac{n}{n \cdot \bar{e} + b} = \underline{\underline{\frac{1}{\bar{e}}}}$$

$$\lim_{n\to\infty} Var[Gamma(\theta; a+n, \sum_{i=1}^{n} e_i + b)] = \lim_{n\to\infty} \ \frac{a+n}{(\sum_{i=1}^{n} e_i + b)^2}$$

$$= \frac{a+n}{\sum_{i=1}^{n} e_i + b} \cdot \frac{1}{\sum_{i=1}^{n} e_i + b} = \frac{1}{\bar{e}} \cdot 0 = \underline{\underline{0}}$$

Again, it is proven that the asymptotic posterior collapses onto a single point, independently of the two initial parameters

### 4.1.3 Convergence of any initial Dirichlet priors on multinomial parameters

**Example:** After studying the fruit flies for an eternity, Adam and Eve have garnered quite the appetite for fruit. They continue to a magical tree which generates different kind of fruit, one at a time. Upon plucking a fruit, a new one appears immediately. Of the selection of pomegranates, peaches, figs, and many more, they wonder how likely each kind of fruit is to appear. Again, they vehemently disagree on how likely each fruit will appear. So while eating many fruit, they keep track of how many fruits of each kind have appeared upon successively plucking them.

**Model:** The prior belief is given by the continuous multivariate distribution $Dirichlet(A)$ s.t $A = \{a_1, a_2, ..., a_k\}$ where $a_i > 0$. The Dirichlet distribution is simply a generalisation of the Beta distribution where the support may be extended to any $k-1$ simplex; A Dirichlet with two categories (k = 2) is the beta distribution. Intuitively, the parameter $a_i$ is called the concentration parameter, which can be understood as how prevalently concentrated the distribution is towards one category. Therefore, $C(\Theta) = f(\Theta; a_1, ..., a_k)$ where $\Theta = \{\theta_1, ..., \theta_k\}$. Each $\theta_i$ ranges over the unit interval. The evidence $E = \{e_1, ..., e_k\}$ is a set of positive integers, whereby each $e_i$ is the sum of successes in each category in i.i.d categorical trials. As such, the size of set $E$ doesn't change but the individual $e_i$ are updated with each observation depending on the observed category. As such $E$ could be written with a subscript denoting the time or how many observations have been made. Consequently then, the likelihood function is the multinomial probability mass function: $\frac{n!}{e_1!e_2!...e_k!} \cdot \theta_1^{e_1}\theta_2^{e_2}\dots\theta_k^{e_k}$, where $n = \sum_{i=1}^{k} e_i$. The PDF of the Dirichlet distribution is given here by: $f(\Theta; A) = \frac{1}{\frac{\prod_{i=1}^{k}\Gamma(a_i)}{\Gamma(a_0)}} \cdot \prod_{i=1}^{k}\theta_i^{a_i-1}$, where $\Gamma(.)$ is the Gamma function again and $a_0 = \sum_{i=1}^{k} a_i$

**Proof of proposition:**

As before, the denominator of the conditionalisation can be omitted as we are only interested in the asymptotic shape of the posterior. Therefore:

$$C(\Theta|E) \propto f(E;n,k,\Theta)\cdot f(\Theta;A) \propto \prod_{i=1}^{k}\theta_i^{e_i}\cdot\prod_{i=1}^{k}\theta_i^{a_i-1} \propto \prod_{i=1}^{k}\theta_i^{e_i+a_i-1}$$
$$\rightarrow C(\Theta|E) = f(\Theta;a_i+e_i)$$

As can be seen, the posterior remains a Dirichlet PDF. Now we check whether the posterior asymptotically collapses onto a single point. As the Dirichlet distribution is multivariate, we can only take the expectation of each axis, or in this case, of each $\theta_i$. Recall that $n = \sum\limits_{i=1}^{k} e_i$, and in the posterior, $a_0^* = \sum\limits_{i=1}^{k} a_i + n$

$$\lim_{n\to\infty} E[C_{\theta_i}] = \lim_{n\to\infty}\frac{a_i+e_i}{\sum\limits_{i=1}^{k}(a_i+e_i)} = \lim_{n\to\infty}\frac{a_i+e_i}{\sum\limits_{i=1}^{k}a_i+n} = \underline{\underline{\frac{e_i}{n}}}$$

$$\lim_{n\to\infty} Var[C_{\theta_i}] = \lim_{n\to\infty}\frac{\frac{a_i+e_i}{a_0^*}\cdot(1-\frac{a_i+e_i}{a_0^*})}{a_0^*+1} = \lim_{n\to\infty}\frac{a_i+e_i}{a_0^{*2}+a_0^*}\cdot\frac{1-\frac{a_i+e_i}{a_0^*}}{a_o^*+1} = 0\cdot ... = \underline{\underline{0}}$$

Which shows that the asymptotic result for each $\theta_i$ is again a single value independent of the initial parameters $A$, and only dependent on the evidence $E$.
The proposition has therefore been proven for the selected conjugate prior-likelihood pairs. Given the pattern in the previous three examples, we could expect the convergence results to hold for all conjugate prior-likelihood pairs. – But even then, it remains uncertain whether the convergence result holds for the thousands of non-conjugate combinations in which Bayesian inference may be applied. As such, more generalised proofs are desired, set in a much richer framework.

## 4.2 Savage 1954

Savage (1954, pg. 46–) demonstrated the first proof for Bayesian convergence in his seminal work *Foundation of Statistics*. The proof was not intended to demonstrate consensus between different agents per se, but instead an asymptotic approach to certainty of the "truth" (Ibid). This certainty arises regardless of the initial prior. In his proof, the hypotheses are a partition of the universal set, composed of every possible state of the world (Savage, 1954, p. 10). To remain consistent with the terminology of Bayesian confirmation theory throughout this paper, we will continue to use the term 'hypothesis' although Savage used the term "event" instead (Savage, pg. 10, 1954). The observations, given by a sequence of outcomes of random variables, take on a finite number of values. These random variables are i.i.d distributed and, so to say, 'governed' according to the "true" hypothesis; As Savage (1954, pg. 47) puts it, it is assumed that given some hypothesis, all observations have the same distribution. By virtue of this assumption, some hypotheses are equivalent, meaning that the distribution of the observed variable does not change given two equivalent hypotheses (Savage, 1954, pg. 48).
The key to the asymptotic convergence lies in the likelihood ratio of alternate hypotheses to the true hypothesis. Essentially, in the long run, there is only one hypothesis which will "obtain" in a mathematically coherent way with the observations of the random variable. Bernardo & Smith (2000, p. 286) provide a neat simplification of Savage's proof. One key difference however is that Savage (1954, pg. 49) does not assume countably additivity as opposed to Bernardo who requires countable additivity to apply the strong law of large numbers.

**Model:** Let $x_j \in \mathbf{x} = \{x_1, x_2, ..., x_\mathbf{n}\}$ be a sequential set of observations which are i.i.d outcomes of a continuous random variable $X$ with a finite mean. This random variable is governed by the true parameter $\theta_t$. All possible parameters $\theta_i$ are finite and "distinguishable" elements of $\Theta = \{\theta_1, \theta_2, ..., \theta_k\}$. The prior is given by $C(\theta_i)$ such that $C(\theta_t) > 0$ and, as is required by probability measures, $\sum_{i=1}^{k} C(\theta_i) = 1$. Specifically here for $\theta_t$, distinguishable is meant in the sense of logarithmic divergence: $\int_X p(x|\theta_t) \cdot \; ln(\frac{p(x|\theta_t)}{p(x|\theta_i)}) \; dx \;\; > \;\; 0 \; \forall t \neq i$. This means the Kullback-Leibler divergence, a measure of distance, is always positive whenever $t \neq i$. This is Bernardo's criterion for distinguishing between equivalent hypotheses.

As noted before, the likelihood function is simply a countably additive unspecified probability function $p(x|\theta)$. Finally, $p(\mathbf{x}|\theta_i)$ can be written out as $\prod_{j=1}^{n} p(x_j|\theta_i)$

**Savage Theorem:**

$$\lim_{n\to\infty} C(\theta_t|\mathbf{x}) = 1 \text{ and } \lim_{n\to\infty} C(\theta_i|\mathbf{x}) = 0 \ \forall t \neq i$$

**Proof:**

$$C(\theta_i|\mathbf{x}) = \frac{p(\mathbf{x}|\theta_i)}{p(\mathbf{x})} \cdot C(\theta_i)$$

$$= \frac{\frac{p(\mathbf{x}|\theta_i)}{p(\mathbf{x}|\theta_t)}}{\frac{p(\mathbf{x})}{p(\mathbf{x}|\theta_t)}} \cdot C(\theta_i) = \frac{R_i}{\sum_{i=1}^{k} C(\theta_i) \cdot R_i} \cdot C(\theta_i) = \frac{e^{S_i + ln(C(\theta_i))}}{\sum_{i=1}^{k} e^{S_i + ln(C(\theta_i))}}$$

where the likelihood ratio is

$$R_i = \frac{p(\mathbf{x}|\theta_i)}{p(\mathbf{x}|\theta_t)} = \prod_{j=1}^{n} \frac{p(x_j|\theta_i)}{p(x_j|\theta_t)} \quad \text{and} \quad S_i = ln(R_i) = \sum_{j=1}^{n} ln\left(\frac{p(x_j|\theta_i)}{p(x_j|\theta_t)}\right)$$

As $X$ is i.i.d with finite mean, we can examine the expectation of $S_i$ using the strong law of large numbers

$$\lim_{n\to\infty} S_i/n = \int_X p(x|\theta_i) \cdot ln\left(\frac{p(x|\theta_i)}{p(x|\theta_t)}\right) dx$$

If $i = t$, then the term in the logarithm equals 0, meaning $S_t \to 0$. If $i \neq t$, then the term in the logarithm explodes negatively, meaning $S_i \to -\infty$. Looking back at the posterior, this yields:

$$\lim_{n\to\infty} C(\theta_t|\mathbf{x}) = \frac{e^{0+ln(C(\theta_t))}}{0 + 0 + ... + e^{0+ln(C(\theta_t))} + ... + 0} = \underline{\underline{1}}$$

$$\forall\ i \neq t \lim_{n\to\infty} C(\theta_i|\mathbf{x}) = \frac{e^{-\infty}}{0 + 0 + ... + e^{0+ln(C(\theta_t))} + ... + 0} = \underline{\underline{0}}$$

## 4.3 Blackwell & Dubins 1962

To further generalize consensus between Bayesian agents, Blackwell & Dubins (1962) rely on Doob's (1953, pg. 319) Martingale Convergence Theorem to prove their seminal merging theorem. To measure consensus, Blackwell and Dubins make the novel decision to use the variational distance between two probability measures; For our objective, these are of course the two agent's credences. Essentially what their proof states is that the agent's credence is a martingale to the information asymptotically gathered, with the same result holding for both agents. As such, it is necessary to first understand the martingale convergence theorem and then subsequently see how the theorem can be applied in Bayesian confirmation theory.

### 4.3.1 Doob's 1953 Martingale Convergence Theorem

Martingales were originally betting strategies hailing from the times of Laplace. Much of the theory on martingales as a stochastic process however came much later with the work of Doob (1953). A later paper by Doob (1971) focuses on the especially pertinent aspects of Martingales for us. Essentially, a martingale $\{X_1, X_2, X_3, ..., X_n\}$ is a sequence of random variables which satisfies the following two conditions:

**1.** $E[|X_n|] < \infty$

**2.** $E[X_{n+1}|X_1, ..., X_n] = X_n$

$X_i$ could be thought of as the price of a stock at each discrete point in time up to some time $n$, and the second condition could imply that the price movement follows a random walk. The real bite behind Martingales comes from Doob's martingale convergence theorem, which requires condition **2.** to be adapted as well as an additional specification.

**1'.** $\sup_{n \in \mathbb{N}} E[|X_n|] < \infty$

**2'.** $E[X_{n+1}|X_1, ..., X_n] \leq X_n$

**1'.** ensures that the martingale is bounded, and **2'.** means that the martingale is a supermartingale; if the sign were $\geq$ instead then it is a submartingale. A clean formulation of Doob's consequent theorem and proof thereof can be taken from the MIT public lecture materials (2013).

**Doob's Martingale Convergence Theorem:**

If conditions **1'.** and **2'.** hold, then $X_\infty = \lim_n X$ exists and is finite, almost surely. The same holds for submartingales.

From the conditions, it may be apparent that the sequence is bounded. However, there is still the possibility of the sequence oscillating. To exclude this possibility, the intuition for the proof depends on Doob's upcrossing lemma. Consider $a, b \in \mathbb{R}$. An upcrossing is when the martingale starts below $a$ and later goes above $b$. Let $U_n[a, b]$ be the number of upcrossings in the martingale until $X_n$. Say that a stock's price series were a supermartingale. A simple strategy would entail buying and selling the stock according to $a, b$ respectively. By the definition of the supermartingale however, there is no reliable strategy that results in a positive gain forever. Therefore, while $U_n[a, b]$ is understandably non-decreasing, there are a finite number of upcrossings almost surely in the infinite martingale sequence, implying that asymptotically the supremum and infimum of the sequence are equal: There is convergence. Essentially, the martingale convergence theorem is the stochastic analogue to the monotone convergence theorem. For further details, consult Durett (2019, pg. 205). The key point now is to have a random variable, e.g an agents belief $C$, be martingale to a monotonically increasing set of information: $C_n = E[C|\mathfrak{F}_n]$ (Doob, 1971, pg. 454; pg. 457). Almost surely then, the random variable will converge onto a finite value. Essentially according to this model, "the game man plays with nature as he learns more and more is fair." (Doob, 1971, pg. 454). The martingale property is even claimed by some as a core principle of rational learning. As Huttegger (2013, pg. 414, 420) states: "[Principle R3: ] An agent's current degree of belief in event $A$ given that her anticipated future degree of belief $P_f(A) = r$ should be equal to $r$ with probability 1, whenever the event $P_f(A) = r$ has positive probability. [...] Principle R3 is, in general, nothing but the martingale property." With the martingale properties in mind, we can finally tackle the most seminal proof.

### 4.3.2 Merging of opinions with increasing information

The original framework of Blackwell & Dubins (1962) is very technically dense, making use of product probability spaces. For better accessibility, we will use the framework of Kalai & Lehrer (1994). Succinct explanations of the merging proof have been provided multiple times (Huttegger, 2015; Nielsen & Stewart, 2018; 2019; Zaffora Blando, 2022). The aim here is detail the proof comprehensively accessible and be mindful of the epistemic prerequisites.

**Example:** Adam and Eve are suddenly thrust into our current world. Being deeply unfamiliar but incredibly curious about the state of this world, they consider different propositional hypotheses with different likelihoods. While they disagree initially on how likely which hypotheses are, will they eventually agree after making infinite observations? To start off simple, they take the closest Swiss Franc coin and start tossing it.

**Model:** Consider a probability space triplet composed of $(\Omega, \mathfrak{F}, c)$. $\Omega$ is the sample space, which can be interpreted as the set of possible worlds which could obtain. One element of this set could include a world where the neighbor's dog is a Labrador, the president of the U.S is Jeb Bush etc., and another element is the world where the neighbor's dog is a Poodle instead. To make the model more manageable, let us consider the set of all possible infinite sequences of coin tosses: $\Omega = \{HHTH..., THHTHH..., \ ...\}$, where $\omega \in \Omega$ is the one sequence which obtains. $\omega$ can be thought of as the one real world. $\mathfrak{F}$ is the sigma algebra of subsets of $\Omega$. The countable elements $h_i \in \mathfrak{F}$ can be understood as events or propositions — hypotheses. For example, any propositions about the outcome of the first or the 62nd coin toss are in $\mathfrak{F}$, but also propositions like limiting properties such as the ratio of heads to tails.
To complete the probability space, $c$ and $\mu$ are countably additive probability measures on the measurable space $(\Omega, \mathfrak{F})$. Most crucially, $c$ and $\mu$ are mutually absolutely continuous: $c \ll \mu$, meaning the agents never observe anything which both held as initially implausible.
Let $\mathfrak{I}_1, \mathfrak{I}_2, ...$ (I in the Fraktur font) be an information sequence for a measurable space. Each element $\mathfrak{I}_n$ is a finite partition of $\Omega$, and each subsequent element $\mathfrak{I}_{n+1}$ refines the previous element. $\mathfrak{I}_n$ can be interpreted as the possible informational dis-

tinctions an agent can make about the world at time $n$. For example, upon the first coin toss, $\Omega$ is partitioned into all sequences starting with heads and all sequences starting with tails. Of course only one subset of $\mathfrak{I}_n$ can obtain, so the agent learns $E_n(\omega) \in \mathfrak{I}_n$. $E_n(\omega)$ can be thought of as the evidence that the agent has at time $n$. Returning to the example, this means the agent learns one of the two subsets created by the first partition.

Now let $\mathfrak{F}_n$ be the sigma algebra generated by $\mathfrak{I}_n$, so $\sigma(\mathfrak{I}_n) = \mathfrak{F}_n$. As $\mathfrak{I}_n$ is refined by $\mathfrak{I}_{n+1}$, a sequential filtration of the sigma algebras ensues: $\mathfrak{F}_n \subseteq \mathfrak{F}_{n+1}$. This entails the crucial notion that the potential learning of the agent always increases. The sigma algebras of the filtration are sub-sigma algebras of $\mathfrak{F}$. The filtration is complete if $lim\mathfrak{F}_n = \mathfrak{F}_\infty = \mathfrak{F}$, which the smallest sigma algebra that contains all of the countable sub-sigma algebras. The information collected now captures all propositions of interest. The agent now conditionalizes the probabilities of any hypotheses: $c_n(\omega) = c(h_i|\mathfrak{F}_n)(\omega) = c(h_i|E_n(\omega))$. These are all *regular* conditional probabilities by virtue of assumptions made about the measurable space.

Finally, we use the variational distance to measure merging between the two probability measures:

$$d(c, \mu) = \sup_{\forall i} |c(h_i) - \mu(h_i)|$$

**Blackwell & Dubins Theorem:**

Let $c$ and $\mu$ be conditionalizing probability measures on $(\Omega, \mathfrak{F})$. If $c \ll \mu$, then $c$ and $\mu$ merge pointwise with $c$ probability 1, i.e

$$c(\{\omega \in \Omega : \lim_{n\to\infty} d(c_n(\omega), \mu_n(\omega)) = 0\}) = 1$$

**Proof sketch:**

The entire proof can very roughly be boiled down to the fact that asymptotically, the agents will observe everything considered possible and can therefore simply evaluate if $\omega \in h_i$. To formalize this result, we need some more structure.

Given that information sequences are generated by finite partitions of $\Omega$, and each

$E_n(\omega)$ has positive probability, we can write the conditionalization in the usual form:

$$c(h_i|\mathfrak{F}_n)(\omega) = \frac{c(h_i \cap E_n(\omega))}{c(E_n(\omega))}$$

How can we be sure that a function like $c(h_i \cap E_n(\omega))$ exists? Here, the Radon-Nikodym Theorem guarantees the existence of this function for any sub-sigma algebra – such as $\mathfrak{F}_n$ – and this function is unique up to sets of $c$-measure 0: $c(h_i \cap E_n(\omega) = \int\limits_{E_n} c(h_i|E_n(\omega))dc$.
Now consider a credence indicator for a given hypothesis $C_h : \Omega \to \{0,1\}$ as a $\mathfrak{F}_n$-measurable random variable. Very informally, this could be interpreted as the approval of a hypothesis by staunch and uninformed agent, or perhaps the rounded credence of an agent. As per the model setup, the conditional expectation $E[C_h|\mathfrak{F}_n]$ of this random variable is bounded (Cf. Ash & Doléans-Dade, 2000, pg. 225). Now consider, by virtue of the tower property of iterated expectations: $E[\ E[C_h|\mathfrak{F}_{n+1}]|\mathfrak{F}_n] = E[C_h|\mathfrak{F}_n]$. Finally, it is rather intuitive that as more potential information becomes available, the hypotheses which don't conform with the obtained information become more unlikely: $E[C_h|\mathfrak{F}_{n+1}] \leq E[C_h|\mathfrak{F}_n]$ if $\omega \notin h$. Conversely, The inequality $\geq$ holds if $\omega \in h$. Ergo, we have supermartingales and submartingales respectively! Finally, consider the indicator function $\forall h \in \mathfrak{F}$:

$$I(\omega; h) = \begin{cases} 1 & \text{if } \omega \in h \\ 0 & \text{if } \omega \notin h \end{cases}$$

Now we have $\lim_{n\to\infty} E[C_h|\mathfrak{F}_n] = E[C_h|\mathfrak{F}_\infty] = E[C_h|\mathfrak{F}] = I(\omega; h)$, so asymptotically the agent will know which world has obtained, and which hypothesis is confirmed by this world.
A necessary technical precaution here is that to ensure clean conditionalization, all conditionalized measures $c(\cdot|E_n(\omega))$ are regular conditional probabilities for all $\omega$. Note also that the convergence obtains almost everywhere $\omega$. Accordingly, if two measures $c$ and $\mu$ are mutually absolutely continuous and conditionalize on the same evidence, then $\lim_{n\to\infty} d(c_n, \mu_n) = 0$ a.e $\omega$.
All following merging theorems will build on Blackwell & Dubins (1962). Hence, further detailed proof sketches are omitted to prevent unnecessary ink spillage.

## 4.4 Schervish & Seidenfeld 1990

While the previous result is undeniably impressive, it remains questionable whether an agent can confidently pick one probability measure and run with it until the end of time. Perhaps an agent would like a selection of probability measures. Now another agent also picks measures from this same selection, or perhaps the second agent even has their own set of measures. Can these two agents be expected to reach a consensus despite their 'shaky' credences? Schervish & Seidenfeld(1990) examine this problem, and Stewart & Nielsen (2019) provide accessible guidance and expand thereupon for their own theorem.

**Example:** Adam and Eve have become exhausted of their studies, but still wish to continue their work on confirming various hypotheses about this new world. As such, they ask for help from some angels. Adam chooses some angels for his work, and Eve selects some other angels. The angels aren't sure which hypotheses are true, and they all disagree on how likely each hypothesis obtains, but at least they are all in mutual agreement as to what hypotheses are plausible. The two groups of angels then all examine the same evidence while Adam and Eve rest. Adam and Eve consult their group of angels regularly. Will Adam and Eve eventually agree on all hypotheses?

**Model and Results:**

Schervish and Seidenfeld (1990, pg. 337) propose a "closed, convex set of probabilities all mutually absolutely continuous, and generated by finitely many of its extreme points". What could a set of probabilities $C$ like this look like? A simple version of such a set could include the probability measures $\mu$, $q$, and $a \cdot \mu + (1-a) \cdot q$ where $a \in [0, 1]$. Visually, this would be a line, and the extreme points of the line are $\mu$ and $q$. Now, this can be expanded to any simplex with finite extreme points – it is crucial that $C$ contains finite extreme points. Therefore in summary, the elements of $C$ can be thought of as differently weighted $a_1 ... a_k$ averages of the unique measures in it: $\sum_{i=1}^{k} a_i = 1$. Finally, the measures in the set $C$ are all mutually absolutely continuous. Conditionalizing for each measure and subsequent regular conditional probability in $C$ works as before.

**Schervish & Seidenfeld Theorem:** For a closed, convex set $C$ of mutually absolutely continuous probability measures, generated by finite extreme points, almost surely the elements of $C$ will merge uniformly (Stewart & Nielsen 2019, Theorem 2), i.e

$$\lim_{n\to\infty} \sup_{\mu,q\in C} d(\mu(\cdot|E_n(\omega)), q(\cdot|E_n(\omega)) = 0 \text{ a.e } \Omega$$

As can be seen in the theorem, the merging result simply needs to hold for the extreme points of the convex set, as the variational distance between all elements of $C$ is bounded by just two extreme points. All things considered, the Blackwell & Dubins Theorem takes care of these two measures. Now what is additionally interesting is how this allows for the omission of dynamic coherence of the two agents. Consider an original convex set $C$ with the assumptions before, and then take any two subsets $C_1, C_2 \subseteq C$. Assign $C_1, C_2$ to two statically coherent agents, i.e these agents will randomly pick probabilities from their assigned sets but will never themselves conditionalize on $E_n(\omega)$. The two agents will still reach a consensus because their sets are doing the conditionalizing work for them. Moving on, the merging results can still be generalized even further.

## 4.5 Uncertain Evidence

So far all merging results have relied upon concrete evidence that is clearly available to the conditionalizing agents. In line with our illustrative examples, van Fraassen (1980, pg. 169) claims that "experience speaks with the voice of an angel". Perhaps so, but it might be hard to understand this angel. Returning to Probability Kinematics/Jeffrey Conditionalization (Jeffrey, 1965), the assumption of certain evidence can be relaxed as well, and merging may still be possible (Huttegger, 2015). Further, this result is incorporated with the previous Schervish & Seidenfeld Theorem (Stewart & Nielsen, 2019).

**Example:**

Adam and Eve cannot access a rare coin in a vault but want to resolve their differing initial hypotheses about it. They ask an angel to flip the coin and relay the results. Due to the distance, they struggle to hear the angel clearly, creating uncertainty about the flip's outcome. Will Adam and Eve still be able to reach a consensus?

### 4.5.1 Huttegger 2015

Information is obtained and priors are conditioned as before. Recall that the sequence of refined partitions $\mathfrak{E}_1, \mathfrak{E}_2$ is composed of elements $e \in \mathfrak{E}_n$. Then for some prior probability measure $c$ the Jeffrey conditionalization looks as follows (Huttegger, 2015, pg. 13):

$$c_n(h_i) = \sum_{e \in \mathfrak{E}_n} c(h_i|e) \cdot c_n(e)$$

Again, if for one element of $\mathfrak{E}_n$ it holds that $c_n(e) = 1$ then we have classic conditionalization again. Now consider the notion of hard and soft Jeffrey shifts (Joyce, 2010, pg. 34; Huttegger, 2015, pg. 15, 18). A hard Jeffrey shift occurs when values of $c_n(e)$ for all $e$ are set, regardless of the prior; The hard shifts are ephemeral. This means that each agent is provided with the same distribution across each iterative piece of evidence, i.e $c_n(e) = \mu_n(e)$. For the example, this means that the agents need to agree for each iteration how likely each potential evidence datum was. To illustrate this, Huttegger (2015, pg. 15) supposes that both agent use the same "mechanical observer", which exactly displays the probability of each evidence datum for an iteration. Juxtapose this with with soft Jeffrey shifts, where the prior is relevant in informing $c_n(e)$, and this prior might have already been informed by previous evidence. As an illustrative example, consider an experienced birdwatcher and a novice hypothesize about the bird species in the area. Both faintly hear a bird go "*coo-woo-woo*". While hearing the same sound, the expert is nearly certain that this came from a morning dove! To provide a concrete mathematical example of hard and soft Jeffrey shifts respectively, consider a partition of $\mathfrak{E}_1$ in three elements (Cf. Huttegger, 2015, pg. 18):

$$c_2(e_1) = 0.1, \; c_2(e_2) = 0.6, \; c_2(e_3) = 0.3$$

versus

$$c_2(e_1) = 4 \cdot c_1(e_1), \; c_2(e_2) = 2 \cdot c_1(e_2), \; c_2(e_3) = 8 \cdot c_1(e_3)$$

Which can, given coincidentally a fitting prior, lead to the same results. Huttegger proves (2015, pg. 30) that only for hard Jeffrey shifts, merging occurs almost surely. Two additional technical specifications need to be made however. First, the sequence of measures $c_n$ is *uniformly* absolutely continuous to $c$. This condition simply ensures that absolute continuity does not get lost in the limit (Huttegger, 2015, pg. 14). Further, the sequence of hard probabilities of the evidence $c_n(e)$ need to form a martingale, formalized as condition $M'$ (Huttegger, 2015, pg. 23). Formally, this means for all elements $G \in \mathfrak{F}$ and all $m \geq n$: $\int_G c_{m+1} dc = \int_G c_m dc$. This can be thought of in the sense that the hard probabilities of the evidence do not have some sort of dynamic structure which the agent can take advantage of. In the classic case, we simply have an uninteresting martingale. If the previous two conditions hold:
**Huttegger Theorem**: Consider a random sequence of martingale probability measures $c_n, \mu_n$ ,which are uniformly absolutely continuous to $c, \mu$, on $(\Omega, \mathfrak{F}_n)$. Let $c_n(e) = \mu_n(e)$. If $c \ll \mu$, then $\lim_{n\to\infty} d(c_n(h_i), \mu_n(h_i)) = 0$ almost surely $\forall h_i \in \mathfrak{F}$.
Unsurprisingly, Huttegger (2015, pg. 21) shows that under soft Jeffrey shifts, credences are not almost surely going to merge. However, Huttegger (pg. 22) does show that under similar martingale conditions, credences converge to some definite limit despite soft Jeffrey shifts.

### 4.5.2 Nielsen & Stewart 2019

Now, consider the above merging result under hard Jeffrey shifts for multiple measures in a convex set according to Schervish & Seidenfeld (1990). Stewart and Nielsen (2019, pg. 244) expand the merging result of hard Jeffrey shifts to convex sets. If the assumptions for the convex set hold – finite extreme points of mutually absolutely continuous probability measures – and the assumptions for the probability kinematics hold as well – uniformly absolutely continuous probabilities which holds in the limit – then merging is guaranteed almost surely.

**Stewart & Nielsen Theorem:** Consider a closed convex set $C$ of a finite number of mutually absolutely continuous probability measure. Each measure is conditioned according to the assumptions of the Huttegger Theorem. Then for $c, \mu \in C$, $\lim_{n\to\infty} d(c_n(h_i), \mu_n(h_i)) = 0$ almost surely $\forall h_i \in \mathfrak{F}$

With the Stewart & Nielsen Theorem, uncertain evidence as well as sets or probability measures can be considered for merging.

## 4.6 Formal language and Algorithmic Randomness

### 4.6.1 Gaifman & Snir 1982

The previous theorems in merging have been restricted to frameworks steeped in probability and measure theory. This and the following theorem do not anymore directly expand upon the previous, but rather add more framework flavour. It was the aim of Gaifman & Snir (1982) to translate the Blackwell & Dubins theorem into rich logical language so as to bring the discussion closer to confirmation theory in analytic philosophy. Synoptic guidance to their result is again given in Earman (1992, pg. 145). Further, Zaffora-Blando (2022) elaborates on notions of algorithmic randomness mentioned in Gaifman & Snir to demonstrate further generalize the Blackwell & Dubins Theorem. As these two results draw heavily from formal language and computability theory, the technical details thereof are beyond the scope of this paper.

Gaifman and Snir (1982, pg. 500) start with a first order language for arithmetic $\mathscr{L}_0$. This language contains names for all integers, and can then describe all sorts of relations across the natural numbers using quantifiers – e.g '$\exists$' – and logical connectives – 'e.g $\wedge$'. Then by adding empirical predicates and empirical function symbols the language $\mathscr{L}$ is created; This means that $\mathscr{L}$ allows for the formulation of empirically testable hypotheses. A *model* is a specific interpretation of the empirical symbols of $\mathscr{L}$, so to say a specific 'world'. Crucially, $\mathscr{L}_0$ is the 'common denominator' of all models. $Mod_L$ is the set of all models for $\mathscr{L}$. As an example to illustrate this framework, the empirical predicate $P(z)$ might state that the $z$th coin flip turns out as 'heads', which is true for some worlds and not so for others. The one real world is $\omega \in Mod_L$. Sentences $\phi$ or $\psi$ of $\mathscr{L}$ – such as the previous coin flip predicate – are true or false depending on the world $\omega$. This is formally captured by *mod*$(\phi) \equiv \omega \in Mod_L : \phi$ *is true in* $\omega$ , which gives the set of all models where $\phi$ is

true (pg. 503). Define $\phi^\omega$ as $\phi$ or $\neg\phi$ if $\omega \in mod(\phi)$ or $\omega \notin mod(\phi)$ respectively; it evaluates the truth of sentence $\phi$. Finally, the evidence is given by $\Phi = \phi_1, \phi_2...$, which separates $Mod_L$, which basically means that the evidence is empirically distinguishable.
In this framework, some probability measure $Pr()$ is defined for sentences $\phi$ of the language $\mathscr{L}$, so that conditional probabilities can be expressed as $Pr(\psi|\phi)$. From the families of sets of $mod(\phi)$ we can form a sigma algebra $\mathscr{F}$. We can now work with a probability space $(Mod_L, \mathscr{F}, Pr)$. With the framework set, we can now make use of the Blackwell Dubins Theorem to develop the following theorem (Earman, 1992, pg. 146).

**Gaifman and Snir Theorem:** Consider the sentences $\phi_i$ which separate $Mod_L$. Then it holds for any two mutually absolutely continuous measures $Pr, Pr'$ that they will merge almost surely for any sentence $\psi$ of $\mathscr{L}$, i.e

$$\lim_{n\to\infty} \sup_{\psi} \left| Pr(\psi| \bigcup_{i\leq n} \phi_i^\omega) - Pr'(\psi| \bigcup_{i\leq n} \phi_i^\omega) \right| = 0$$

This theorem could be expanded to the Schervish & Seidenfeld Theorem (Earman, 1992, pg. 147). – Perhaps it could even be expanded to the Huttegger and Stewart & Nielsen Theorems.

### 4.6.2 Zaffora-Blando 2022

As Gaifman and Snir (1982, pg. 496) note: "In all proposals randomness is defined as the satisfaction of a certain class of properties that have probability 1. [...] A *notion of randomness* is nothing other than the satisfaction of all sentences of probability 1 which belong to a certain class". Zaffora-Blando further examines these notions of randomness to further examine the Blackwell Dubins theorem. As she states (2022, pg. 3): "Each algorithmic randomness notion corresponds [...] to a precise class of effectively specifiable global regularities". These different notions of randomness include Martin-Löf randomness and Schnorr randomness. Agreement between agents about the randomness underpinning the data stream can imply absolute continuity of their probability measures, and therefore also merging (Zaffora-Blando, 2022, pg. 25). What Zaffora-Blando therefore manages is to give technical form to an understanding of the Humean uniformity of nature – providing a beautiful link to the qualitative notion which gave impetus to confirmation theory. As she (2022, pg. 26) concludes: "Bayesian agents [...] who nonetheless agree on which data streams are algorithmically random, may be thought of as having compatible inductive assumptions about the uniformity of nature." To go into the exact technicalities of these notions of randomness is beyond the scope of this paper.

Besides the mentioned merging theorems, one more by Sterkenburg & De Heide (2021) will be touched upon in chapter 6 within the discussion about mutual absolute continuity. Otherwise, we have now covered merging theorems spanning roughly 70 years of Bayesian Confirmation theory. In ascending complexity, this has covered the prior-likelihood convergence, Savage's theorem (1954), Doob's martingale convergence theorem (1953), Blackwell & Dubin's Theorem (1962), Schervish & Seidenfeld's Theorem (1990), Huttegger's Theorem (2015), Nielsen & Stewart's Theorem (2019), and further Gaifman & Snir (1982) and Zaffora-Blando (2022).

# 5 Epistemic Onion

Having summarized a selection of Bayesian merging theorems, we can now make a comprehensive account of the epistemic assumptions required by two agents in order for them to reach consensus. With the previous theorems in mind, we can return to the epistemic onion constructed earlier in chapter 3. Despite a large set of initial assumptions, the subsequent theorems are able to formally generalize the older theorems. The assumptions listed are not to be interpreted as perfectly individually necessary for any modeled outcome, but rather as jointly sufficient conditions that intermingle and overlap at times. As such, it is more so taking a route down a long road starting at a core and taking specific branches.

## 5.1 Assumptions by Theorem

### 5.1.1 Conjugate Prior-Likelihood

The first merging theorem comes in the conjugate prior-likelihood setting. As this is the most basic merging scenario, we need all of the previously listed assumptions, with a high emphasis on the probability coordination principle to get a specific form of the likelihood function in line with the distribution of the i.i.d evidence. Additionally, the prior is also tightly constrained to a specific probability density function. Finally, the hypotheses are of a tame statistical type – They would state: “The parameter of the distribution of the evidence is $\theta_i$”. This gives us the following most salient requirements for merging:

**Conjugate prior-likelihood assumptions:**

- The evidence is i.i.d distributed according to some specific distribution given the hypothesis

- The credences range over statistical hypotheses
- The likelihood function is a specific probability function conjugate to the prior

### 5.1.2 Savage

Savages Theorem allows us to generalize the likelihood function as well as the prior. Interestingly, Savage does not make use of countable additivity in his original proof as he uses only the Weak Law of Large Numbers, and as such the credences asymptotically only become $\epsilon > 0$ close. Accordingly, Savage (1954, pg. 68) states: "The conclusion of the personalistic view is not that evidence brings holders of different opinions to the same opinions, but rather to similar opinions." We have adapted the proof to get merging, which required us to add countable additivity. As such, we can adapt the previous salient assumptions.

**Savage assumptions:**

- The evidence is i.i.d distributed according to some distribution given the hypothesis
- The credences range over statistical hypotheses

### 5.1.3 Blackwell & Dubins

With Blackwell & Dubins, we are exiting the strict statistical corset. The assumption of i.i.d trials of a specific distributions is no longer necessary. The probability coordination principle is severely weakened to the extent where it is questionable if it even still applies, as not even a common likelihood function is required. With the more technical setup, the assumptions however shift as well. A mild assumption is that there are regular conditional probabilities.

**Blackwell & Dubins assumptions:**

- For some measurable space $(\Omega, \mathfrak{F})$, the two select credences are mutually absolutely continuous

### 5.1.4 Schervish & Seidenfeld

The Schervish & Seidenfeld theorem allows for the agents to simply pick measures from a convex set. To reiterate, the closed convex set needs to be generated by a finite number of extreme points. As the agents are simply randomly picking from the convex set of conditionalizing measures, they technically don't need to conditionalize themselves. However, this just kicks the conditionalization down the road, and as such, conditionalization is not dropped as an assumption.

**Schervish & Seidenfeld assumptions:**

- For some measurable space $(\Omega, \mathfrak{F})$, the mutually absolutely continuous credences can be arbitrarily picked from a convex set.

### 5.1.5 Huttegger

The Huttegger theorem allows us to drop the assumption of definite evidence by leveraging hard Jeffrey shifts. An important requirement for these hard Jeffrey shifts is however that the martingale condition $M'$ holds for the hard probabilities.

**Huttegger assumptions**

- For some measurable space $(\Omega, \mathfrak{F})$, the mutually absolutely continuous credences conditionalize on uncertain evidence

### 5.1.6 Nielsen & Stewart

Finally, the Nielsen & Stewart theorem allows us to extend the Huttegger theorem to a convex set as per the Schervish & Seidenfeld theorem.

**Huttegger assumptions**

- For some measurable space $(\Omega, \mathfrak{F})$, the mutually absolutely continuous credences which conditionalize on uncertain evidence can be arbitrarily picked from a convex set.

## 5.2 Comparison

Comparing the sufficient assumptions of each subsequent merging proof, we receive the following table. For the exact meaning of the assumptions in the left column, please refer to the earlier sections. Red circles mean the assumption can be omitted.

**Table 1: Epistemic Requirements**

| Assumption | Conjugate prior-likelihood | Savage | Blackwell & Dubins | Schervish & Seidenfeld | Huttegger | Nielsen & Stewart |
|---|---|---|---|---|---|---|
| Separability | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Propositions | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Observational hypotheses | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Boolean logic | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Formal language | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Propositional logic | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Propositional attitude | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Doxastic attitude | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Rationality | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Belief ranking | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Real number assignment | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Subjective probabilities | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Probability space | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Countable additivity | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Common iter. evidence | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Uniformity of nature | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Martingale | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Regular cond. prob. | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Jeffrey conditionalization | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Equally dogmatic | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| Convex set of measures | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 |
| One extreme point | 🟢 | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 |
| Probability 1 evidence | 🟢 | 🟢 | 🟢 | 🟢 | 🔴 | 🔴 |
| Prob. coord. principle | 🟢 | 🟢 | 🔴 | 🔴 | 🔴 | 🔴 |
| Statistical hypotheses | 🟢 | 🟢 | 🔴 | 🔴 | 🔴 | 🔴 |
| I.i.d evidence | 🟢 | 🟢 | 🔴 | 🔴 | 🔴 | 🔴 |
| Conjugate prior-likelihood | 🟢 | 🔴 | 🔴 | 🔴 | 🔴 | 🔴 |

The Gaifman & Snir theorem has not been included as it translates all other findings into specific formal language. As such, the sufficient assumptions of Gaifman & Snir have been assumed beforehand as they more so ensure that the other results can fit into a language framework. The details of the Zaffora-Blando theorem has not been included as the technicalities of algorithmic randomness are beyond the scope of this paper, however the uniformity of nature remains an important principle.
It becomes apparent on the table that while numerous assumptions need to hold for any merger of beliefs, a decent number of assumptions have been omitted over time. It is hard to estimate how many more assumptions can feasible be omitted while maintaining a coherent and precise framework for 'agreement'; To simply make use of the total variation distance, a large swathe of probabilistic assumptions already apply. Another point to be mentioned is that the school of Bayesian confirmation theory is steeped in an analytic tradition. This makes precise adjustments and fixes of the model readily implementable, however this approach also causes a great deal of reductionism when modelling actual human learning processes. What doesn't help in this regard is the "orgulity" (Belot, 2013) of some members of this school of thought. For our present purposes – to consider highly ideal and analyzable agents for disagreement – Bayesian confirmation theory has provided fruitful insights.
What seems to be a sticking point for all merging results is the requirement of mutual absolute continuity. This assumption will be tackled by means of a new model in the following penultimate chapter.

# 6 Creeping

## 6.1 Previous Discussion

The requirement of mutual absolute continuity has proven so essential to the Blackwell & Dubins Theorem that Kalai & Lehrer (1994, pg 81) prove that merging also conversely implies continuity. As such, this requirement has become deeply etched into the conditions for consensus. Upon reconsidering classic Bayesian conditionalization, there seems to be no orthodox way that a hypotheses with zero initial credence will ever gain any credence. Therefore, unorthodox methods are explored in an attempt to weaken this requirement.

To tie into the academic discussion, consider one of the key challenges of Bayesian confirmation theory: The Problem of New Theories (Strevens, 2017, pg. 129; Earman, 1992, pg. 133). As the name suggests, there seems to be no straightforward means for a Bayesian agent to embrace novel hypotheses previously unconsidered. As a Bayesian agent must assign credence one to the union of all disjoint hypotheses, the hypotheses beyond this set are deemed as impossible. To give this problem more form, consider the two logical omniscience assumptions LO1 and LO2 previously touched upon (Earman, 1992, pg. 121). LO1 states that all logical truths of the language used are readily apparent to the agent and its credences. LO2 states that the agent is aware of all possible hypotheses relevant. To illustrate this we will return to the coin example. LO1 suggests that if an agent observes a single heads outcome, it is clear that the coin doesn't have an extreme bias resulting exclusively in tails outcomes. LO2 suggests that the agent is aware that for a coin the true weighting parameter lies anywhere on the unit interval. To "humanize" the Bayesian agent, LO2 seems to be a good assumption to attack. An agent who fails LO2 could fail to assign positive probability to the true hypotheses. In order to 'shift' the set of hypotheses deemed plausible to include the true hypotheses, the agent must somehow adopt a new subjective probability function. As Earman (1992, pg. 133) states, this new function is typically "not derived [...] by any straightforward conditionalization process".

To this end, consider first the motivations a rational agent might have to embrace new hypotheses. While most motivations for subjective probabilities go back to gambling (Ramsey, pg. 176), consider specifically the motivation of an agent to attain certainty of some true hypothesis. Rosenkrantz (1981) provides a formal motivation for Bayesian conditionalization on the basis of minimizing the loss generated by giving credence to wrong hypotheses. Earman (1992, pg. 44) provides a succinct summary of the motivation. Consider a number of hypotheses and one true hypothesis $h_i, h_t \in H$. The inaccuracy or 'untruthfulness' loss function of the agent's belief are then expressed by the credences $c_i$ assigned to each respective hypotheses

$$I(c; \theta_i) = c_1^2 + c_2^2 ... + (1 - c_t)^2 + c_{t+1}^2 ... + c_k$$

Previously, truth has been a conduit for certainty of some specific hypotheses, and then all agents with differing priors will merge on that point. Now, consider further the truth as some center of gravity which 'pulls' the agent's credences towards it from any direction.

One common approach to the problem of new theories is the inclusion of a 'catch-all' hypothesis, a concept coined by Shimony (1970, pg. 200). For a set of considered hypotheses $H_c$, the catch-all hypotheses $\bar{H}_c$ is the complement thereof, so in formal terms $H = H_c \cup \bar{H}_c$. The catch-all hypothesis can also be considered a reservoir of credence which the agent can assign to any new probability, hence why Strevens (2017, pg. 131) calls it the "reserve hypothesis". A principled method for devising new methods from this reservoir is via the "shaving off" principle (Earman, 1992, pg. 196). As new hypotheses are introduced, their credence is gained by shaving off from the reserve hypothesis, implying that over time this well of fresh hypotheses dries up. Wenmackers and Romeijn (2016) translated this idea into a statistical framework, dubbing it "open-minded" Bayesianism. The newest iteration of an open-minded Bayesian model comes from Sterkeburg and De Heide (2021), which modifies the model to ensure the convergence to truth, and by extension therefore

also the merger of opinions.
We would like to propose another principled response to the problem of new theories which – in our opinion – truly maintains the spirit of a more humanized Bayesianism.

## 6.2 Cut and Creep

As Earman (1992, pg. 197) already remarked, a feasible solution to the problem of new theories must expand on the usual conditionalization in an novel way. Upon tackling this problem, the requirement for equal dogmatism amongst agents for merging may also be weakened. As such we propose a 'creeping support' model.
Consider the classic coin tossing example with statistical hypotheses relating to the weighting parameter. Two agents have credences $C$ and $M$ over these hypotheses. Now, consider further a weakening of LO2, so the agents do not know that the true parameter $\theta_t$ lies in the unit interval. We say weakening because for merging in the creeping model, we still require the probability coordination principle so that physical likelihoods can be used. The physical likelihood calculation needs $\theta_i \in [0, 1]$ for sensible probabilities, so in some sense, the agents do know by virtue of the likelihood function that the true parameter must lie in the unit interval. In an equivalent sense to not knowing, the agent could also simply deem certain ranges of the unit interval as implausible a priori. As such in this setup, we may receive two agents where $C(\theta_i) \neq M(\theta_i)\ \forall\ i$. Figure 5 shows the priors chosen for this setup. As seen on Figure 5, the inner borders of the priors are here given by 0.4 and 0.6 respectively. Consider $\theta_t = 0.5$, which is considered implausible a priori by both agents. By classically conditioning upon the evidence as seen in the first conjugate prior example in chapter 4, the two posteriors will converge to the border of the prior. Figure 6 shows the development of the credences after 500 iterations under classic conditionalization.

**Figure 5:**

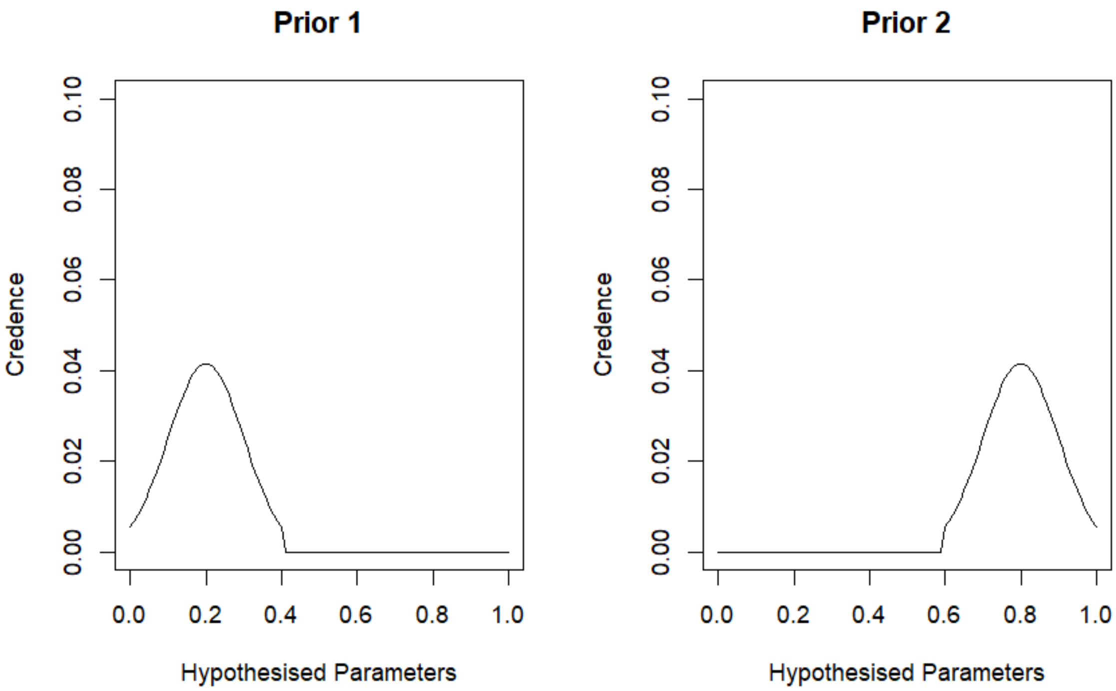
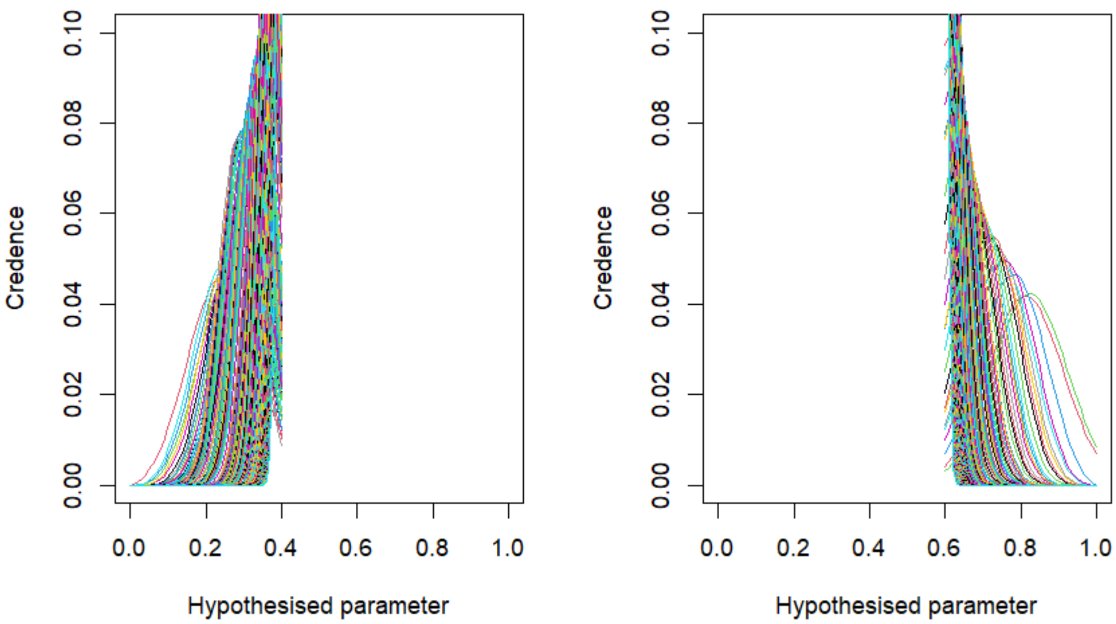



**Figure 6:**



A person confronted with his credences slamming against the borders of the prior would surely be disposed to curiosity in that direction. The posteriors appears as if gravitating towards the true parameter. Then, the creeping kicks in: For some subjective threshold parameter $\gamma$, any hypotheses below this credence are discarded and considered implausible. – But to maintain conformity with probability calculus, where should the 'cut' probability go? The agent needs to determine the directional tendency of the posterior, and will then essentially swap the probability from one side to another. On the unit interval, this direction can easily be given by $-1$ for left and 1 for right. As such, the support of the agent's credence creeps iteratively in the direction of the convincing hypotheses. Hopefully, both credences will then converge to the true parameter, which leads to merging again. This creeping model does not make use of some catch-all hypothesis. By simply slowly shifting the support in the convincing direction, the creeping model remains true to a humanized version of logical omniscience.

While merging between agents is a welcomed effect of the creeping model, the individual and rational impetus behind creeping can be justified by the inaccuracy loss function from earlier. By creeping, an agent may become certain of the true hypothesis. What remains trickier to justify is the directional tendency and the threshold parameter $\gamma$. For statistical hypotheses in the coin tossing example, a variety of statistical metrics could be employed to determine the 'direction' of a distribution. The threshold $\gamma$ however might have to be solely subjectively justified. One way to justify $\gamma$ might be to assume the complement of the Lockean thesis. Inspired by John Lock, Foley (1992, pg. 111) characterizes the Lockean thesis as follows:

> "To say that we believe a proposition is just to say that we are sufficiently confident of its truth for our attitude to be one of belief. Then it is epistemically rational for us to believe a proposition just in case it is epistemically rational for us to have sufficiently degree of confidence in it."

A complement to this sufficient degree of confidence could substantiate $\gamma$. Considering the original context of subjective probabilities, $\gamma$ could relate to some betting stake considered negligible to the agent. Outside monetary gambling, $\gamma$ can be interpreted as how unsubstantiated a hypothesis needs to be for an agent to reject it, making it a kind of measure of conservatism or steadfastness in the face of evidence. In the following application, the directional tendency and $\gamma$ were chosen in a way to facilitate convergence to truth and by extension merging.

## 6.3 Application and Discussion

The final creeping model is implemented in R (See appendix script 'Creeping Model.R') and assumes two agents in the setup described before. A seed is set for reproducibility, but the model has been developed without one. The agents share a common memory of the binary evidence outcomes, which are stored in a vector. They classically conditionalize on the evidence. The direction of the posterior is measured by taking the median index of the posterior, and then checking which side of the median has more probability mass. 1 or -1 is the direction of confidence based on if more mass is to the right or left of the median index respectively. Using the negative skewness of the distribution has not worked favourably for merging. Then, the posterior is checked for any credences not in the direction of confidence below $\gamma$; these credences are to be cut and redistributed to the other side. The first slot where they may be assigned are the credences below $\gamma$ in the direction of confidence. This is done so that there are no sudden cuts or valleys in the distribution in case the direction of confidence also has credences below $\gamma$. If there are not enough available slots for the below threshold credences to be redistributed to, then they are rescaled via a custom function in order to fit the amount of slots. So for example, if credences $0.5, 0.1, 0.1, 0.1$ had to be redistributed on two available slots, then they would be rescaled to $0.6, 0.2$. Finally, after redistribution towards the direction of confidence, the cut credences are assigned probability 0 to maintain a sum of 1 of the distribution. One small fix implemented is that the credences for the extreme hypotheses where $\theta_1 = 0$ and $\theta_k = 1$ are deductively deduced. This means that as soon as for example a 1 appears in the evidence, $\theta_1$ is assigned zero as it was just proven that a 1 can appear, as we do not want this hypothesis to creep back in. To detect any potential biases in the model, the directional confidence of each agent is stored for each iteration in order to check that not one direction is somehow systematically prefered because of an accidental asymmetry in the model. Considering all this, the distributions are plotted with each iteration.

Regarding parameters of the model, there are aptly called meta-parameters, prior parameters, and learning parameters, all of which are incredibly relevant to the behaviour of the creeping.

Regarding the first, the number of hypotheses is assigned, where a high number allows for rough approximations of a PDF. The number of periods or iterations is assigned, with a high number aiming to manifest asymptotic behaviour. Finally, a true parameter for the binary outcomes is set. For balanced results with decent computational efficiency, 100 hypotheses were chosen, 800 periods, and a true parameter of 0.5 for symmetry.

The prior parameters determine the mean and standard deviation of a normal distribution, which will then generate the a normal distribution prior of the agents. This gives us the priors displayed earlier in Figure 5.

Finally, the learning parameters cause the most volatility. The threshold parameter $\gamma$ is selected as the mean of the credences above zero times some factor. A higher factor causes more erratic creeping, and a lower factor makes the agent more conservative. $\gamma$ can also be dynamic, meaning that it is always changes every iteration as a function of the posterior. Next, the learning occurring via the likelihood function can be 'accelerated' or 'braked'. This is done with a learning rate $lr$ and a modulation exponent $root$. Consider the numerator of Bayes theorem, then these parameters are applied as follows: $lr \cdot likelihood^{root} \cdot prior$. The modulation exponent is especially useful if set $< 1$ in order to prevent hasty and excessive credence in wrong hypotheses. In our result, $lr$ is set to 1 and $root$ is set to 0.05. Finally, there is a *wait* parameter which allows for intermittent creeping. For example, in our case *wait* is set to 10, then only every 10th iteration will the agent creep along the parameter space. Combined with conservative updating, this allows for more controlled and paced creeping. Considering all the mentioned parameters, the following values seen on table 2 were used for the presented simulation:

**Table 2: Selected Parameters**

| True Parameter | Number of Hypotheses | Periods | Prior Mean | Prior SD | Initial $\gamma$ | Learning Rate | Root | Wait |
|---|---|---|---|---|---|---|---|---|
| 0.5 | 100 | 800 | 0.2, 0.8 | 0.1 | mean prior | 1 | 0.05 | 10 |

Applying all these parameters, the following learning development seen on Figure 7 ensues over 800 iterations.
The final posterior narrowly concentrates around the true parameter of 0.5, as seen on Figure 8. The horizontal red line indexes the hypotheses with the highest credence. For the given seed and periods, this resulted in 0.48 and 0.49 for agent 1 and 2 respectively.

As a proof of concept, the creeping model can as such be considered a success and may serve as a new unorthodox model to explore new theories for agents failing to initially consider all physically possible hypotheses. Many drawbacks and problems need to be mentioned however.
First of all, the creeping model is incredibly wild and unpredictable. Despite the parameters that induce 'conservatism', the posteriors may still settle on a wrong hypothesis. In the future, a systematic analysis of the parameters could be undertaken to inspect what parameter combinations result is best convergence to truth results. On a related point, the model has been coded with an approach that favours understandability of the coded model over computational efficiency. Many functions could be precoded, and then vectorized. This would also allow for better approximations of continuous functions, and faster testing of the parameters. On a more conceptual basis, the direction of confidence metric and redistribution algorithm could be adapted. More complex and established metrics might give a better metric for the 'direction' of the distribution. This also relates to the next point, namely how the model is to be expanded to higher order simplexes. How the direction of confidence and the redistribution is determined could have significant impacts on how creeping manifests when multiple parameters are to be estimated, like for example weights of a die. Finally, a direction of confidence remains highly ambiguous for non-statistical hypotheses.
Despite the numerous areas for improvement, the creeping model has proven its efficacy in allowing for merger of opinions through convergence to truth despite a violation of initial mutual absolute continuity. Perhaps some day, a formal proof can be derived that agents who declare themselves always open to creep will be certain of merging.

**Figure 7:**

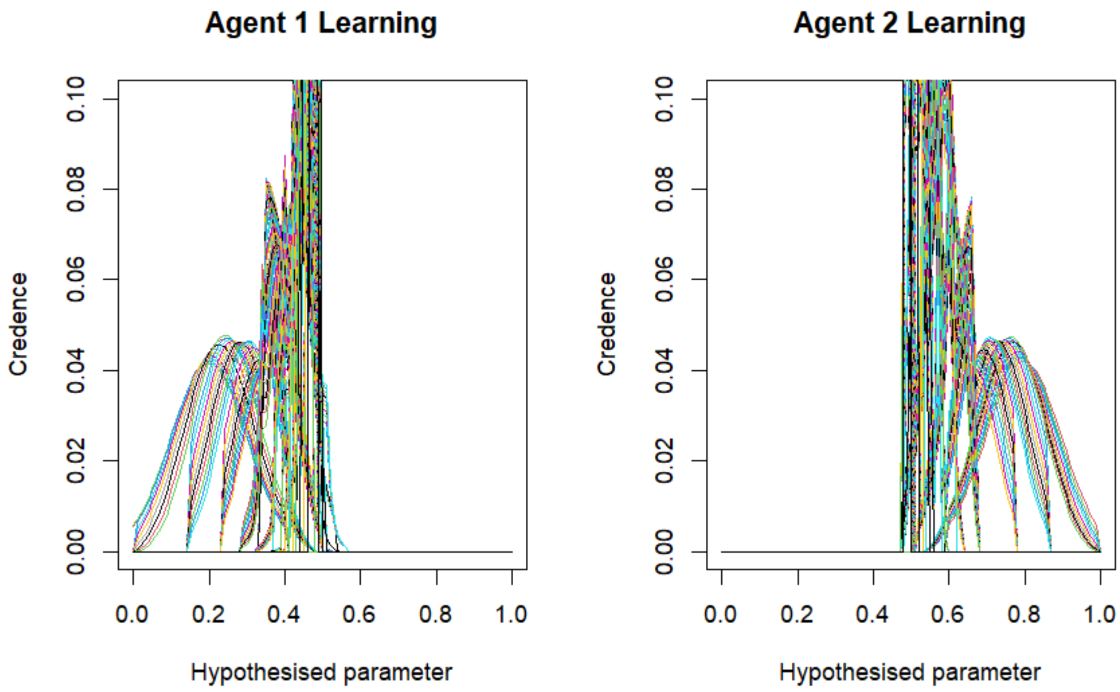
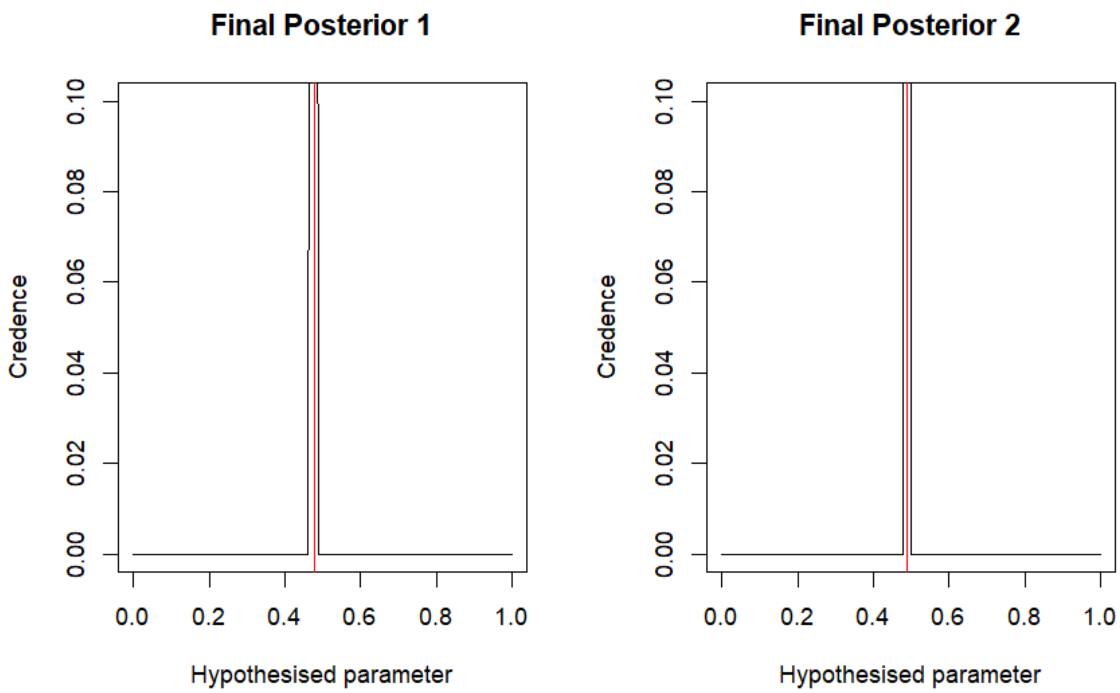



**Figure 8:**



# 7 Conclusions

Arriving at the end of this paper, we will now summarize the findings and supply an outlook for future research in this domain. This paper has presented a jointly sufficient list of assumptions for the Bayesian merging theorems, specific asymptotic proofs for the conjugate prior setting, assumptions which can be omitted by virtue of the field's development, and a model to tackle equal initial dogmatism.

Returning to the original proposition of TOTAL, it becomes clear that the resolution of disagreement between rational agents upon learning the same evidence is not remotely guaranteed. By specifying an epistemic agent imbued with an account of rationality sufficient for Bayesian inference, a plethora of analytic assumptions need to apply for consensus to obtain asymptotically. Diving into the increasingly complex models, from Savage (1954) to Nielsen & Stewart (2019), we could omit multiple assumptions demanded by the originally strictly statistical Bayesianism. For a common formal language, given the same uncertain evidence, equally dogmatic agents can rely on a set of a priori disagreeing measures with subjective probabilities across hypotheses – and merging will obtain almost surely. This is undeniably impressive, however there remain many more underlying assumptions. The novel creeping model allows merging even for agents which are not equally dogmatic initially, given of course a simple setting. As such, taking the sufficient and overlapping requirements listed in table 2 of chapter 5 and then giving the agents infinite time, it becomes clear that consensus is not something we as mere humans can almost surely expect within our lifetimes.

A glaring first critique obviously lies in the infinite horizon of merging results. For finite horizons, $\epsilon$-close weak merging (Lehrer & Smorodinsky, 1996) offers applicable results. Calculating rates of convergence to certainty for the complex models could also help in this regard. Regarding the creeping model, a formal parameter optimization could greatly benefit merging results, and an expansion to more complex

settings will demand a reinterpreted application of underlying principles. Furthermore, while Bayesianism is the most studied theory of confirmation, alternatives do exist which might reduce the pool of assumptions (Cf. Bandyopadhyay & Forster, 2011, pg. 394). Also, there is a specific account of rationality that runs throughout the merging theorems discussed and the analytic school thereof; tackling a deeper understanding of rationality could perhaps result in broader merging theorems.
With the merging results seeming to rest upon such fickle assumptions, a liberal society of individuals agreeing on everything might seem like a fleeting hope. However, within a society we also form our beliefs not just based on the type of evidence discussed in this paper, but also by influence of our peers. By allowing for interactions between agents, agreement and merging by extension could become far more interesting (Cf. Jehle & Fitelson, 2009; Weatherall & O'Connor, 2020). Further, core beliefs in principles prior to evidence such as the uniformity of nature have proven necessary for merging (Blando, 2022). Therefore, for future research to generalize merging, perhaps we need not entirely sacrifice the analytical corset which allows for such rigorous theorems. Moreover, the epistemic agent should be analyzed within a broader organic context of social life that includes heritage, values, and peers, with both learning and behavior therein influencing the agent's worldview. In such a humanized context, an understanding of consensus might yet reap the solutions needed to cure the contemporary ailments of destructive disagreement.

# Bibliography

[1] Abramowitz, M., & Stegun, I. A. (Eds.). (1972). "Gamma function and related functions" in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (p. 258). New York: Dover Publications. ISBN 978-0-486-61272-0. Specifically, see 6.2 Beta Function.

[2] Ash, R. B., & Doléans-Dade, C. A. (2000). *Probability and Measure Theory.* Academic Press.

[3] Bandyopadhyay, P. S., & Forster, M. R. (2011). Philosophy of statistics. In *Handbook of the Philosophy of Science* (Vol. 7). Amsterdam: Elsevier. Retrieved from `https://www.researchgate.net/publication/296788330_Philosophy_of_Statistics`

[4] Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 53, 370–418. `http://doi.org/10.1098/rstl.1763.0053`

[5] Belot, G. (2013). Bayesian Orgulity. *Philosophy of Science*, Vol. 80, No. 4 (October 2013), pp. 483-503. Retrieved from `https://www.jstor.org/stable/10.1086/673249`

[6] Bernardo, J. M. (2000). *Bayesian Theory.* Retrieved from `https://statisticalsupportandresearch.files.wordpress.com/2019/03/josc3a9-m.-bernardo-adrian-f.-m.-smith-bayesian-theory-wiley-1994.pdf`

[7] Blackwell, D., & Dubins, L. (1962). Merging of Opinions with Increasing Information. *The Annals of Mathematical Statistics*, Vol. 33, No. 3, pp. 882-886. Retrieved from `https://www.jstor.org/stable/2237864?seq=5`

[8] Blando, Z. (2022). Bayesian merging of opinions and algorithmic randomness. Retrieved from https://static1.squarespace.com/static/5d4e174e652fc20001100877/t/6500dd5fc96758438a64ef0d/1694555487427/Bayesian+merging+of+opinions+and+algorithmic+randomness.pdf

[9] Carnap, R. (1950). *Logical Foundations of Probability*. Chicago: University of Chicago Press.

[10] Conee, E. (2010). Rational Disagreement Defended. *Oxford*, 2010; online edn, Oxford Academic, 1 Sept. 2010. Retrieved from https://doi.org/10.1093/acprof:oso/9780199226078.003.0005

[11] Davis, P. J. (1972). "Gamma function and related functions" in Abramowitz, M., & Stegun, I. A. (Eds.), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (p. 258). New York: Dover Publications. ISBN 978-0-486-61272-0. Specifically, see 6.2 Beta Function.

[12] De Finetti, B. (1975) [2017]. *Theory of Probability*. Retrieved from https://onlinelibrary.wiley.com/doi/book/10.1002/9781119286387

[13] Doob, J. L. (1953)[1990]. *Stochastic Processes*. John Wiley & Sons, New York.

[14] Durrett, R. (2019). *Probability: Theory and examples*. Retrieved from https://services.math.duke.edu/~rtd/PTE/PTE5_011119.pdf

[15] Earman, J. (1992). *Bayes or Bust?* MIT Press.

[16] Foley, R. (1992). The Epistemology of Belief and the Epistemology of Degrees of Belief. *American Philosophical Quarterly*, 29(2), 111–124. Retrieved from http://www.jstor.org/stable/20014406

[17] Freedman, D., & Diaconis, P. (1986). On the consistency of Bayes Estimates. *Annals of Statistics*, 14(1): 1-26. DOI: 10.1214/aos/1176349830. Retrieved from https://projecteuclid.org/journals/annals-of-statistics/volume-14/issue-1/On-the-Consistency-of-Bayes-Estimates/10.1214/aos/1176349830.full

[18] Gaifman, H., & Snir, M. (1982). Probabilities over rich Languages. University of Illinois Urbana-Champaign. Retrieved from https://snir.cs.illinois.

edu/listed/J5.pdf

[19] Hawthorne, J. (2011). Confirmation. In *Handbook of the Philosophy of Science, Philosophy of Statistics*, (Vol. 7, pp. 333-389).

[20] Hume, D. (1748)[1999]. *Enquiry concerning human understanding.* In T. Beauchamp (Ed.). Oxford University Press. Retrieved from http://139.59.56.236/bitstream/123456789/489/1/An%20Enquiry%20concerning%20Human%20Understanding.pdf

[21] Huttegger, S. (2013). In Defense of Reflection. *Philosophy of Science*, Vol. 80, No. 3 (July 2013), pp. 413-433. Retrieved from https://www.jstor.org/stable/10.1086/671427?seq=2

[22] Huttegger, S. (2015). Merging of Opinions and Probability Kinematics. *The Review of Symbolic Logic*, Volume 0, Number 0, Month 2009. Retrieved from https://bpb-us-e2.wpmucdn.com/faculty.sites.uci.edu/dist/c/190/files/2011/03/MergingRSLRevision3.pdf

[23] James, W. (1897). *The Will to Believe.* In *The Will to Believe and Other Essays in Popular Philosophy* (pp. 1–15). New York: Longmans, Green, and Co.

[24] Jaynes, E. T. (1968). Prior Probabilities. *IEEE Transactions On Systems Science and Cybernetics*, vol. sec-4, no. 3, pp. 227-241. Retrieved from https://bayes.wustl.edu/etj/articles/prior.pdf

[25] Jeffreys, H. (1945). An invariant form for the prior probability. *The Royal Society.* Retrieved from https://royalsocietypublishing.org/doi/10.1098/rspa.1946.0056

[26] Jeffreys, H. (1961). *Theory of Probability.* Retrieved from https://mathscinet.ams.org/mathscinet/relay-station?mr=0187257

[27] Jeffrey, R. C. (1965)[1983]. *The Logic of Decision.* New York: McGraw-Hill. Third revised edition. Chicago: University of Chicago Press, 1983. Retrieved from https://fitelson.org/piksi/the_logic_of_decision.pdf

[28] Jehle, G., & Fitelson, B. (2009). What is the equal weight view? *EPISTEME*, 2009. Retrieved from https://joelvelasco.net/teaching/3865/jehlefitelson09-equalweight.pdf

[29] Joyce, J. M. (2010). The development of subjective Bayesianism. In D. M. Gabbay, S. Hartmann, & J. Woods (Eds.), *Handbook of the History of Logic, Volume 10: Inductive Logic* (pp. 415–476). Elsevier. Retrieved from https://websites.umich.edu/~jjoyce/papers/dsb.pdf

[30] Kalai, E., & Lehrer, E. (1993). Rational Learning leads to Nash Equilibrium. *Econometrica*, Vol. 61, No. 5 (Sep., 1993), pp. 1019-1045. Retrieved from https://www.jstor.org/stable/2951492?seq=1

[31] Kalai, E., & Lehrer, E. (1994). Weak and Strong merging of Opinions. *Journal of Mathematical Economics*, Volume 23, Issue 1, January 1994, Pages 73-86. Retrieved from https://www.sciencedirect.com/science/article/pii/030440689490037X

[32] Keynes, J. M. (1921)[2020]. *A Treatise on Probability*. Alpha Editions.

[33] Kolmogorov, A. (1933)[2013]. *Foundations of the theory of probability*. Martino Publishing.

[34] Kvanvig, J. (2014). *Rational Disagreement*. Retrieved from https://doi.org/10.1093/acprof:oso/9780198716419.003.0005

[35] Kvanvig, J. (2023). Confirmation. Washington University in St. Louis. Retrieved from https://wustl.app.box.com/s/tmg5rb3ng9jp22mt6fjspfslak9qofmf

[36] Lehrer, E., & Smorodinsky, R. (1996). Merging and Learning. *Lecture Notes-Monograph Series*, Vol. 30, Statistics, Probability and Game Theory: Papers in Honor of David Blackwell (1996), pp. 147-168. Retrieved from https://www.jstor.org/stable/4355944?seq=1

[37] Levi, I. (1980). *The Enterprise of Knowledge*. Cambridge: The MIT Press.

[38] Lindley, D. V. (2000). The Philosophy of Statistics. *Journal of the Royal Statistical Society. Series D (The Statistician)*, Vol. 49, No. 3, pp. 293-337. Retrieved from https://www.jstor.org/stable/2681060?seq=1

[39] MIT (2013). Martingale convergence theorem. *MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 6.265/15.070J Fall 2013, Lecture 11-Additional material 10/9/2013.* Retrieved from https://ocw.mit.

edu/courses/15-070j-advanced-stochastic-processes-fall-2013/66b6c8fdb52304e3777ce8286beaaf7d_MIT15_070JF13_Lec11Add.pdf

[40] Nielsen, M., & Stewart, C. (2018). Persistent Disagreement and Polarisation in a Bayesian Setting. *The British Journal for the Philosophy of Science*, Volume 72, Number 1. Retrieved from https://www.journals.uchicago.edu/doi/abs/10.1093/bjps/axy056?journalCode=bjps

[41] Norton, J. (2011). Challenges to Bayesian Confirmation Theory. In *Handbook of Philosophy of Statistics* (Vol. 7, pp. 391-439).

[42] Ramsey, F. P. (1926). "Truth and Probability", in Ramsey, F. P. (1931). *The Foundations of Mathematics and other Logical Essays*, Ch. VII (pp. 156-198). Edited by R.B. Braithwaite. London: Kegan, Paul, Trench, Trubner & Co., Ltd; New York: Harcourt, Brace and Company. Retrieved from https://fitelson.org/probability/ramsey.pdf

[43] Raiffa, H., & Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Retrieved from https://gwern.net/doc/statistics/decision/1961-raiffa-appliedstatisticaldecisiontheory.pdf

[44] Raju, C.K. (2011). Probability in Ancient India. In *Handbook of Philosophy of Statistics* (Vol. 7, pp. 1175-1196). Elsevier.

[45] Rosenkrantz, R. (1981). *Foundations and Applications of Inductive Probability*. Ridgeview.

[46] Savage, L. J. (1954) [1972]. *The Foundations of Statistics*. Retrieved from https://gwern.net/doc/statistics/decision/1972-savage-foundationsofstatistics.pdf

[47] Schervish, M. J., & Seidenfeld, T. (1990). An Approach to Consensus and Certainty with Increasing Evidence. *Journal of Statistical Planning and Inference*, Volume 25. Retrieved from https://www.cmu.edu/dietrich/philosophy/docs/seidenfeld/Approach%20to%20consensus%20and%20certainty.pdf

[48] Shimony, A. (1970)[1993]. Scientific inference. In *The Search for a Naturalistic World View* (Vol. II, pp. 183–273). Cambridge: Cambridge University Press. Retrieved from https://www.cambridge.org/core/

books/search-for-a-naturalistic-world-view/scientific-inference/BF223D4F6C9ECF057AC0C0234F3AE372

[49] Skyrms, B. (1987). Dynamic Coherence and Probability Kinematics. Retrieved from https://www.jstor.org/stable/pdf/187470.pdf?refreqid=fastly-default%3Aa95cd027f76f680158f94a9025cff920&ab_segments=&origin=&initiator=&acceptTC=1

[50] Sterkenburg, T. F., & De Heide, R. (2021). On the truth-convergence of open-minded Bayesianism. *The Review of Symbolic Logic*, Volume 15, Issue 1. Retrieved from https://philsci-archive.pitt.edu/18786/1/baydyn.pdf

[51] Stewart, C., & Nielsen, M. (2019). Another Approach to Consensus and Maximally Informed Opinions with Increasing Evidence. *Philosophy of Science*, Volume 86, Issue 2. Retrieved from https://www.cambridge.org/core/journals/philosophy-of-science/article/another-approach-to-consensus-and-maximally-informed-opinions-with-increasin
AFEFFFA9619E0E8309901E6CD2BC4451

[52] Strevens, M. (2017). *Notes on Bayesian confirmation theory.* Retrieved from http://www.strevens.org/bct/

[53] Titelbaum, M. (2022). *Fundamentals of Bayesian epistemology 1.* Oxford University Press.

[54] Tu, S. (2014). Dirichlet-Multinomial conjugate prior form. Retrieved from https://stephentu.github.io/writeups/dirichlet-conjugate-prior.pdf

[55] Van Fraassen, B. C. (1980). Rational belief and probability kinematics. *Philosophy of Science*, 47, 165–187.

[56] Weatherall, J. O. (2020). Endogenous Epistemic Factionalization. *Synthese.* Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3304109

[57] Wenmackers, S., & Romeijn, J.W. (2016). New theory about old evidence. *Synthese*, 193, 1225–1250. Retrieved from https://doi.org/10.1007/s11229-014-0632-x

# 9 Appendix

## 9.1 R scripts

### 9.1.1 Script: Bayesian Simulations.R

```
library(tidyverse)
library(gbutils)
library(extraDistr)
library(moments)
rm(list = ls())
set.seed(1)
par(mfrow = c(1, 2))
####
# Bayesian Updating
####

p_values <- seq(0,1, 0.001) #hypothesis range

# Prior creedence for theta (n assumed known), discretized
prior <- dunif(p_values, 0, 1)

prior <- prior/sum(prior) #normalize

ylim = 0.01
plot(prior, type = "l", ylim = c(0,ylim))

#theta true value
theta = 0.3


#Bayesian Learning(pg. strevens 2017, pg. 53)-----
```

```
#Basic case, only one agent

periods <- 100
for (i in 1:periods){

e <- rbinom(1 , 1 , theta)#draws,range, p-parameter

post <- dbinom(e,1,p_values ) * prior #physical likelihoods according to pdf * prior

post <- post/sum(post) #divided by law of total probability

lines(post, col = (1+i)) #plot posterior

prior <- post #post is next periods prior

#Sys.sleep(0.8) #for animation purpose
}


plot(prior ~ p_values, type = "l", ylim = c(0,ylim))
abline( v=  ( p_values[which(prior == max(prior)) ]) , col="red"  )
p_values[which(prior == max(prior)) ] #mode


####
####
########## Two Agent case, different initial priors, same hypothesis range -----
rm(list = ls())
dev.off()
n = 1000 #number of possible hypotheses

p_values <- seq(0,1, 1/n) #hypothesis range

prior.param1 <- 0.2 #prior 1 mode
prior1 <- dbeta(p_values, 1, 2) #prior should not be 0 anywhere!
prior1[p_values < 0] <- 0
prior1[p_values > 1] <- 0
prior1 <- prior1/sum(prior1)
```

```
prior.param2 <- 0.7 #prior 1 mode
prior2 <- dbeta(p_values, 2, 1) #prior should not be 0 anywhere!
prior2[p_values < 0] <- 0
prior2[p_values > 1] <- 0
prior2 <- prior2/sum(prior2)

#plot priors next to each other
par(mfrow = c(1,2))

ylim = 0.01
par(mfg = c(1, 1))
plot(prior1 ~ p_values, type = "l", ylim = c(0,ylim), xlab = "Hypothesised parameter",
     ylab = "Credence", main = "Agent 1 Prior")

par(mfg = c(1, 2))
plot(prior2 ~ p_values, type = "l", ylim = c(0,ylim), xlab = "Hypothesised parameter",
     ylab = "Credence", main = "Agent 2 Prior")

#theta true value
theta = 0.3

#Bayesian updating

periods <- 800

for (i in 1:periods){

  e <- rbinom(1 , 1 , theta)#draws,range, p-parameter

  post1 <- dbinom(e,1,p_values ) * prior1 #physical likelihoods according to pdf * prior
  post1 <- post1/sum(post1) #divided by law of total probability


  post2 <- dbinom(e,1,p_values ) * prior2 #physical likelihoods according to pdf * prior
  post2 <- post2/sum(post2) #divided by law of total probability

  #Adding priors to graphs
  par(mfg = c(1, 1))
  lines(post1 ~ p_values, col = 1+i )
```

```
  par(mfg = c(1, 2))
  lines(post2 ~ p_values, col = 1+i )

  prior1 <- post1 #post is next periods prior
  prior2 <- post2 #post is next periods prior

  #Sys.sleep(0.8)

}
dev.off()
par(mfrow = c(1,2))

par(mfg = c(1, 1))
plot(prior1 ~ p_values, type = "l", ylim = c(0,ylim*2), xlab = "Hypothesised parameter",
     ylab = "Credence", main = "Agent 1 Final Posterior")
abline( v=  ( p_values[which(prior1 == max(prior1)) ]) , col="red"  )
p_values[which(prior1 == max(prior1)) ] #mode

par(mfg = c(1, 2))
plot(prior2 ~ p_values, type = "l", ylim = c(0,ylim*2), xlab = "Hypothesised parameter",
     ylab = "Credence", main = "Agent 2 Final Posterior")
abline( v=  ( p_values[which(prior2 == max(prior2)) ]) , col="red"  )
p_values[which(prior2 == max(prior2)) ] #mode


####
###### border cases - exclusion of true hypothesis by both agents

rm(list = ls())
dev.off()
n = 100 #number of possible hypotheses

lbound1 <- 0
ubound1 <- 0.4
p_values1 <- seq(lbound1 ,ubound1 , 1/n) #hypothesis range for agent 1

lbound2 <- 0.6
ubound2 <- 1
p_values2 <- seq(lbound2 ,ubound2 , 1/n) #hypothesis range for agent 2
```

```
prior.param1 <- 0.2 #prior 1 mode
prior1 <- dnorm(p_values1, prior.param1, sd = 0.1) #prior should not be 0 anywhere!
prior1[p_values1 < lbound1] <- 0
prior1[p_values1 > ubound1] <- 0
prior1 <- prior1/sum(prior1)


prior.param2 <- 0.8 #prior 1 mode
prior2 <- dnorm(p_values2, prior.param2, sd = 0.1) #prior should not be 0 anywhere!
prior2[p_values2 < lbound2] <- 0
prior2[p_values2 > ubound2] <- 0
prior2 <- prior2/sum(prior2)

#plot priors next to each other
par(mfrow = c(1,2))

ylim = 10/n
par(mfg = c(1, 1))

plot(prior1 ~ p_values1, type = "l", ylim = c(0,ylim),
     xlab = "Hypothesised Parameters", ylab = "Credence", main = "Prior 1")

par(mfg = c(1, 2))
plot(prior2 ~ p_values2, type = "l", ylim = c(0,ylim),
     xlab = "Hypothesised Parameters", ylab = "Credence", main = "Prior 2")

#theta true value
theta = 0.5


#Bayesian updating

periods <- 500

plot(0, type = "n", xlim = c(0, 1), ylim = c(0,ylim), xlab = "Hypothesised parameter",
     ylab = "Credence")
plot(0, type = "n", xlim = c(0, 1), ylim = c(0,ylim), xlab = "Hypothesised parameter",
     ylab = "Credence")
```

```
for (i in 1:periods){

  e <- rbinom(1 , 1 , theta)#draws,range, p-parameter

  ph.like1 <- dbinom(e,1,p_values1 )
  ph.like1 <- ifelse(is.na(ph.like1), 0, ph.like1)
  post1 <- ph.like1 * prior1 #physical likelihoods according to pdf * prior
  post1 <- post1/sum(post1) #divided by law of total probability


  ph.like2 <- dbinom(e,1,p_values2 )
  ph.like2 <- ifelse(is.na(ph.like2), 0, ph.like2)
  post2 <- ph.like2 * prior2 #physical likelihoods according to pdf * prior
  post2 <- post2/sum(post2) #divided by law of total probability

  #Adding priors to graphs
  par(mfg = c(1, 1))
  lines(post1 ~ p_values1, col = 1+i ) #VISUAL BUG HERE, DOES NOT PLOT PROPERLY

  par(mfg = c(1, 2))
  lines(post2 ~ p_values2, col = 1+i )

  prior1 <- post1 #post is next periods prior
  prior2 <- post2 #post is next periods prior

  #Sys.sleep(0.8)

}

dev.off()
par(mfrow = c(1,2))

par(mfg = c(1, 1))
plot(prior1 ~ p_values1, type = "l", ylim = c(0,ylim*2))
abline( v=  ( p_values1[which(prior1 == max(prior1)) ]) , col="red"  )
p_values1[which(prior1 == max(prior1)) ] #mode
```

```
par(mfg = c(1, 2))
plot(prior2 ~ p_values2, type = "l", ylim = c(0,ylim*2))
abline( v=  ( p_values2[which(prior2 == max(prior2)) ]) , col="red"  )
p_values2[which(prior2 == max(prior2)) ] #mode




####
## Wrong conditionalisation; Violation of physical likelihood by agent 2 ----

rm(list = ls())
dev.off()
n = 1000 #number of possible hypotheses

lbound1 <- 0
ubound1 <- 1
p_values1 <- seq(lbound1 ,ubound1 , 1/n) #hypothesis range for agent 1

lbound2 <- 0
ubound2 <- 1
p_values2 <- seq(lbound2 ,ubound2 , 1/n) #hypothesis range for agent 2


prior.param1 <- 0.15 #prior 1 mode
prior1 <- dnorm(p_values1, prior.param1, sd = 0.1) #prior should not be 0 anywhere!
prior1[p_values1 < lbound1] <- 0
prior1[p_values1 > ubound1] <- 0
prior1 <- prior1/sum(prior1)


prior.param2 <- 0.5 #prior 1 mode
prior2 <- dnorm(p_values2, prior.param2, sd = 0.1) #prior should not be 0 anywhere!
prior2[p_values2 < lbound2] <- 0
prior2[p_values2 > ubound2] <- 0
prior2 <- prior2/sum(prior2)

#plot priors next to each other
par(mfrow = c(1,2))
```

```
ylim = 0.01
par(mfg = c(1, 1))
plot(prior1 ~ p_values1, type = "l", ylim = c(0,ylim))

par(mfg = c(1, 2))
plot(prior2 ~ p_values2, type = "l", ylim = c(0,ylim))

#theta true value
theta = 0.3



#Bayesian updating

periods <- 500

for (i in 1:periods){

  e <- rbinom(1 , 1 , theta)#draws,range, p-parameter

  ph.like1 <- dbinom(e,1,p_values1 )
  ph.like1 <- ifelse(is.na(ph.like1), 0, ph.like1)
  post1 <- ph.like1 * prior1 #physical likelihoods according to pdf * prior
  post1 <- post1/sum(post1) #divided by law of total probability


  ph.like2 <- dnbinom(e,1,p_values2 ) #negative Binomial distribution (wrong!!!)
  #Same effect also possible by changing size of outcomes to e.g 2!
  ph.like2 <- ifelse(is.na(ph.like2), 0, ph.like2)
  post2 <- ph.like2 * prior2 #physical likelihoods according to pdf * prior
  post2 <- post2/sum(post2) #divided by law of total probability

  #Adding priors to graphs
  par(mfg = c(1, 1))
  lines(post1 ~ p_values1, col = 1+i ) #VISUAL BUG HERE, DOES NOT PLOT PROPERLY

  par(mfg = c(1, 2))
  lines(post2 ~ p_values2, col = 1+i )

  prior1 <- post1 #post is next periods prior
```

```
  prior2 <- post2 #post is next periods prior

  #Sys.sleep(0.8)

}

dev.off()
par(mfrow = c(1,2))

par(mfg = c(1, 1))
plot(prior1 ~ p_values1, type = "l", ylim = c(0,ylim*2))
abline( v=  ( p_values1[which(prior1 == max(prior1)) ]) , col="red"  )
p_values1[which(prior1 == max(prior1)) ] #mode

par(mfg = c(1, 2))
plot(prior2 ~ p_values2, type = "l", ylim = c(0,ylim*2))
abline( v=  ( p_values2[which(prior2 == max(prior2)) ]) , col="red"  )
p_values2[which(prior2 == max(prior2)) ] #mode
```

### 9.1.2 Script: Creeping Model.R

```
library(tidyverse)
rm(list = ls())
set.seed(1)
par(mfrow = c(1, 2))
dev.off()
###########-
#functions-----
rdtrs <- function(vec, k) { # This function allows for rescaling in case there aren't
                            #enough slots
  result <- numeric(k)

  # Determine the sum for each of the k segments, following a specific logic
  # The first segment takes a different portion than the others
  first_segment_ratio <- sum(vec[1:(length(vec)/k)]) / sum(vec)
  remaining_sum_ratio <- (1 - first_segment_ratio) / (k - 1)

  for (i in 1:k) {
```

```
    if (i == 1) {

      result[i] <- first_segment_ratio * sum(vec)
    } else {

      result[i] <- remaining_sum_ratio * sum(vec)
    }
  }

  return(result)
}

####
#Bayesian Creeping ----

#Meta parameters-----

n = 100 #number of possible hypotheses
periods <- 800 #number of iterations
theta = 0.5 #theta true value

####
####Setting the priors for two agents----
omn.lbound <- 0 #omniscient bounds
omn.ubound <- 1

lbound1 <- 0 #lower bound
ubound1 <- 0.4 #upper bound

p_values1 <- seq(omn.lbound ,omn.ubound, 1/n) #hypothesis range for agent 1

lbound2 <- 0.6
ubound2 <- 1
p_values2 <- seq(omn.lbound,omn.ubound , 1/n) #hypothesis range for agent 2


prior.param1 <- 0.2 #prior 1 mean
prior1 <- dnorm(p_values1, prior.param1, sd = 0.1)
prior1[p_values1 < lbound1] <- 0
prior1[p_values1 > ubound1] <- 0
```

```
prior1 <- prior1/sum(prior1)


prior.param2 <- 0.8 #prior 1 mean
prior2 <- dnorm(p_values2, prior.param2, sd = 0.1)
prior2[p_values2 < lbound2] <- 0
prior2[p_values2 > ubound2] <- 0
prior2 <- prior2/sum(prior2)

#plot priors next to each other
par(mfrow = c(1,2))

ylim = 10/n
par(mfg = c(1, 1))
plot(prior1 ~ p_values1, type = "l", ylim = c(0,ylim),
      xlab = "Hypothesised Parameters", ylab = "Credence", main = "Prior 1")

par(mfg = c(1, 2))
plot(prior2 ~ p_values2, type = "l", ylim = c(0,ylim),
     xlab = "Hypothesised Parameters", ylab = "Credence", main = "Prior 2")


#Learning Parameters-----

g<- mean(prior1[prior1 > 0]) #cut parameter gamma, function of prior

wait <- 10
lr <- 1 #learning rate
root <- 0.05


#memory of evidence
Ev <- vector(mode = "numeric", length = periods)

#extreme hypothesis (these need to be deductively checked)
zero.theta <- NA
one.theta <- NA

plot(0, type = "n", xlim = c(omn.lbound, omn.ubound), ylim = c(0,ylim),
     xlab = "Hypothesised parameter", ylab = "Credence", main = "Agent 1 Learning")
```

```
plot(0, type = "n", xlim = c(omn.lbound, omn.ubound), ylim = c(0,ylim),
     xlab = "Hypothesised parameter", ylab = "Credence", main = "Agent 2 Learning")

dir.hist <- as_tibble( matrix(NA, nrow = n, ncol = 2))  #to keep track of any biases
                                                        #in direction




################-
#################-

for (i in 1:periods){
  #print iteration -----
  print(paste0("iteration ", i))

  Ev[i] <- rbinom(1 , 1 , theta)#binary evidence outcome: draws,range, p-parameter,
                                #saved in memory

  ph.like1 <- dbinom(sum(Ev[1:i]), i, p_values1 )
  ph.like1 <- ifelse(is.na(ph.like1), 0, ph.like1)
  post1 <-  lr*(ph.like1)^(root) * prior1 #physical likelihoods according to
                                          #pdf * prior
  post1 <- post1/sum(post1) #Normalization



  ph.like2 <- dbinom(sum(Ev[1:i]), i,p_values2 )
  ph.like2 <- ifelse(is.na(ph.like2), 0, ph.like2)
  post2 <-    lr*(ph.like2)^(root) * prior2 #physical likelihoods according to
                                            #pdf * prior
  post2 <- post2/sum(post2) #Normalization

  #Adding priors to graphs
  par(mfg = c(1, 1))
  lines(post1 ~ p_values1, col = 1+i )

  par(mfg = c(1, 2))
  lines(post2 ~ p_values2, col = 1+i )


  agents <- list(post1, post2) #putting priors in list so we can loop
```

```
#wait condition -----
if (i %% wait == 0){ #only cutting every wait-th period

  #cut and creep-----
  for(j in 1:length(agents)) {

    #dynamic gamma
    g <- mean(agents[[j]][agents[[j]] > 0] )


    t.replaced <- ( (agents[[j]] < g) & (agents[[j]] != 0) ) #hypotheses with low
                                                              #credence

    n.replaced <- sum( t.replaced ) #how many hypotheses to be replaced

    median.index <-((( max(which(agents[[j]] > 0)) -
                         min(which(agents[[j]] > 0)))/2)
                    + (min(which(agents[[j]] > 0))) )

    #what direction does the agent favour? Maybe neither?
    dir.confidence <- ifelse(
      sum(agents[[j]][1:round(median.index)]) >
        sum(agents[[j]][floor(median.index):length(agents[[j]])]), -1, 1)
    dir.confidence <- ifelse(
      sum(agents[[j]][1:round(median.index)]) ==
        sum(agents[[j]][floor(median.index):length(agents[[j]])]), 0, dir.confidence)

    #dir hist save----
    #dir.hist[i,j] <- dir.confidence #to keep track of the directions chosen


    spots <- (agents[[j]] == 0) #available hypotheses spots to be filled

    if(is.na(zero.theta)){
      zero.theta <- ifelse(Ev[i] == 1, FALSE, NA)} #logical exclusion of
                                                    #extreme hypotheses with memory
    if(!is.na(zero.theta)){
      spots[1] <- zero.theta
    }
```

```
if(is.na(one.theta)){
  one.theta <- ifelse(Ev[i] == 0, FALSE, NA)}#logical exclusion of
                                             #extreme hypotheses with memory
if(!is.na(one.theta)){
  spots[n+1] <- one.theta
}




#creeping
if(  n.replaced > 0  ) {  #Only start creeping if there are any
                          #unsubstantiated hypotheses


  if(dir.confidence == -1) {#want to creep left
    spots[floor(median.index):n+1] <- FALSE #general crawl to the left only
    overlay <- t.replaced#these hypotheses are in good direction but low credence
    overlay[floor(median.index):n+1] <- FALSE
    spots <- spots | overlay

    t.replaced[1:ceiling(median.index)] <- FALSE
    n.replaced <- sum(t.replaced)
    if(n.replaced > 0){



    move.cred <- agents[[j]][t.replaced] #credences of unsubstantiated
                                         #hypotheses in wrong direction


    if( sum(which(spots)) > 0  ){ #to make sure there are available spots

      index.spots <- which(spots)

      if(length(index.spots) < n.replaced){ #rescaling if necessary
        move.cred <- rdtrs(move.cred, length(index.spots)  )
      }

      #creeping left
```

```
      agents[[j]][ rev(index.spots)[1:min(n.replaced,length(index.spots))]] <-
        move.cred + agents[[j]][ rev(index.spots)[1: min(n.replaced,
                                                  length(index.spots))]]

    }
  }

  } else if(dir.confidence == 1) { #want to creep right
    spots[1:ceiling(median.index)] <- FALSE #general crawl to the right only
    overlay <- t.replaced
    overlay[1:ceiling(median.index)] <- FALSE #these hypotheses are in good
                                          #direction but low credence
    spots <- spots | overlay

    t.replaced[floor(median.index):(n+1)] <- FALSE
    n.replaced <- sum(t.replaced)
    if(n.replaced > 0){

    move.cred <- agents[[j]][t.replaced] #credences of unsubstantiated
                                      #hypotheses in wrong direction

    if( sum(which(spots)) > 0  ){ #to make sure there are available spots

      index.spots <- which(spots)

      if(length(index.spots) < n.replaced){ #rescaling if necessary
        move.cred <- rdtrs(move.cred, length(index.spots)  )
      }

      #creeping right
      agents[[j]][ index.spots[
        1: min(n.replaced, length(index.spots)) ] ] <- rev(move.cred) +
        agents[[j]][ index.spots[1: min(n.replaced, length(index.spots)) ] ]

    }
    }

  }
}
```

```
      agents[[j]][t.replaced] <- 0 #deem hypotheses below threshold implausible

    } #end agent creeping loop ----

  }#end of wait: not every period cutting-----

  prior1 <- agents[[1]] #post is next periods prior
  prior2 <- agents[[2]] #post is next periods prior



  #sys sleep -----
  Sys.sleep(0.6) #to animate development


}



mean(dir.hist$V1, na.rm = TRUE) #average direction chosen for prior1
mean(dir.hist$V2, na.rm = TRUE) #average direction chosen for prior2



dev.off()
par(mfrow = c(1,2))

#plotting

par(mfg = c(1, 1))
plot(prior1 ~ p_values1, type = "l", ylim = c(0,ylim),
     xlab = "Hypothesised parameter", ylab = "Credence", main = "Final Posterior 1")
abline( v=  ( p_values1[which(prior1 == max(prior1)) ]) , col="red"  )
p_values1[which(prior1 == max(prior1)) ] #mode

par(mfg = c(1, 2))
plot(prior2 ~ p_values2, type = "l", ylim = c(0,ylim),
     xlab = "Hypothesised parameter", ylab = "Credence", main = "Final Posterior 2")
abline( v=  ( p_values2[which(prior2 == max(prior2)) ]) , col="red"  )
p_values2[which(prior2 == max(prior2)) ] #mode
```

## Declaration of Authorship

I hereby declare

- that I have written this thesis without any help from others and without the use of documents or aids other than those stated above;
- that I have mentioned all the sources used and that I have cited them correctly according to established academic citation rules;
- that I have acquired any immaterial rights to materials I may have used, such as images or graphs, or that I have produced such materials myself;
- that the topic or parts of it are not already the object of any work or examination of another course unless this has been explicitly agreed to with the faculty member in advance and is referred to in the thesis;
- that I will not pass on copies of this work to third parties or publish them without the university's written consent if a direct connection can be established with the University of St.Gallen or its faculty members;
- that I am aware that my work can be electronically checked for plagiarism and that I hereby grant the University of St.Gallen copyright in accordance with the Examination Regulations insofar as this is required for administrative action;
- that I am aware that the university will prosecute any infringement of this declaration of authorship and, in particular, the employment of a ghostwriter, and that any such infringement may result in disciplinary and criminal consequences which may result in my expulsion from the university or my being stripped of my degree.

By uploading this academic term paper, I confirm through my conclusive action that I am submitting the Declaration of Authorship, that I have read and understood it, and that it is true.

St. Gallen
09.03.2024
Noé Kuhn

## On the Usage of Chat GPT

I have used chat gpt for proof-reading purposes, to order alphabetically and clean the Bibliography, and for brainstorming inspiration on topics generally.