A Brief Survey of Formal and Informal Tools for Prediction

Brendan A. Schuetze

Brendan.Schuetze@utexas.edu

Winter 2019

Authors Note (3/26/19): This essay was written for my graduate course in History and Systems of Psychology, which focused on the analysis of different psychological paradigms through the use of an ontology-epistemology-methodology framework. Many of the authorial choices reflected in this paper were made due to course requirements. Nevertheless, I thoroughly enjoyed writing these brief sketches concerning the development of predictive frameworks over time. I hope that my argument will be thought-provoking for those who choose to read further.

Abstract

In this paper, I make use of Paul Meehl's informal-formal dichotomy for understanding the types of predictive frameworks employed throughout the history of the human species. Using this framework, I classify and sketch out three main eras of prediction and the interrelations between them:

1. **Cognition as Prediction (Prehistory):** In the vein of Karl Friston's work on the free energy principle, I understand the human brain as the fundamental predictive engine, upon which all other models have been bootstrapped.

2. **Informal Belief-Based Frameworks (35,000 BCE – 1500 CE):** Thus, I argue that language and complex social systems, such as religion, exist in order to communicate, systematize, and improve the informal accuracy of the predictive brain.

3. **Rise of Formal Predictive Models (1500 CE – Present):** With the ability to offload cognitive processes using language, human expression gave rise mathematical expression of formal predictions, a class of models that has grown increasingly capable since the late 1800s.

Lastly, I predict that future directions in prediction will include the increasingly close relationship between informal and formal models of prediction.

A Brief Survey of Formal and Informal Tools for Prediction

Prognostication, forecasting, projecting, prophesizing, and fortune-telling are all synonyms for the same underlying process – predicting what is to occur in a period of time that has not yet transpired. At a high-level, these words may refer to essentially the same action, however, the connotations associated with these words evince a richer history of prediction. Prognostication, prophesizing and fortune-telling all connote some sort of qualitative and perhaps religious or mystical components to the way in which predictions are made. Forecasting and projection typically imply a reliance on mathematical models.

Correspondingly, I adopt and expand upon Paul Meehl's informal-formal dichotomy (Meehl, 1954), which is often adopted in comparative studies of prediction (e.g., Grove et al., 2000; Grove & Meehl, 1996). Under Meehl's framework, modes of prediction are categorized into two broad categories, "informal" and "formal." Informal prediction refers to the intuitive process by which humans make predictions, whereas formal prediction refers to the use of explicit mathematical models. Informal frameworks have also been denoted by the terms of "clinical," "subjective," "holistic," and "impressionistic," whereas formal models have sometimes been referred to as "mechanistic" and "algorithmic" (Grove & Meehl, 1996).

The prototypical example of an informal predictive model in widespread use today is the process by which doctors traditionally diagnose patients. The diagnosis process is considered informal, because it relies on the integration of information regarding the patient's illness by the clinical practitioner, without the explicit use of mathematical models or statistics. Though informal models have historically dominated clinical practice, this has begun to change since the advent of the twenty-first century. Formal diagnosis frameworks are increasingly being used in medical settings to arrive at a diagnosis by estimating the risks of conditions associated with the presence or absence of different risk factors (Sniderman, D'Agostino, & Pencina, 2015).

Formal and informal models are not mutually exclusive; in fact, they are often used to augment one-another. Today, informal models and formal models coexist in the domains of university admissions (e.g., Kuncel, Klieger, Connelly, & Ones, 2013), academic assessment (e.g., Truesdell & Bath, 1957; Wiggens & Kolen, 1971), psychological assessment (e.g., Westen & Weinberger, 2004), and management (e.g., Lewis & MacKinney, 1961). Seeing as even the most nascent formal predictive models have only existed for a little under four centuries (Howarth, 2001), informal models have been the dominant mode of prediction throughout history, regardless of application (Kuncel, et al., 2013). In fact, some researchers argue that the functional purpose of the brain is to integrate information and engage in continuous prediction (Friston, 2009). Nevertheless, I argue that the burgeoning development of formal predictive models over the last two hundred years has led to the greatest period of upheaval in the landscape of predictive frameworks since the evolution of *homo sapiens sapiens*.

A running theme throughout this paper can be summarized as follows: prediction is a uniquely important human behavior that has been augmented through the use of increasingly-complex theoretical frameworks and mathematical tools. Not only have the *tools and frameworks* changed, but the *practice* of prognostication has also fluctuated through time and within different populations. For example, even within the modern era, the manner which education researchers, medical doctors, governments, and religious figures make predictions varies. And this variation exists between and within these groups; not all education researchers use the same tools or identify with the same constructs.

One fact, however, remains constant. All individuals and groups holding power in modern society – whether political, bureaucratic, or cultural – rely on some framework through which the future can be predicted and thus controlled (van Otterlo, 2014). Throughout the history of prognostication, we will see that the tools and models of prediction are highly influential factors underlying the practice of prediction and the control of society. Predictive frameworks circumscribe the type of questions that are asked, the type of data that is collected and the resulting decisions.

The study of prediction is intertwined with the study of its tools and frameworks. Or, as media-theorist Marshall McLuhan notes, "we shape our tools, and thereafter our tools shape us" (Austen, 2011, n.p.; cf. Rao, 2018). Rather than organizing this paper in relation to key individuals or events, this paper is organized by dominant predictive frameworks in order of first appearance. As such, this paper will advance from the cognition as prediction model to religious frameworks, to early pre-scientific (i.e., mechanistic paradigms), to the advanced statistical and machine-learning techniques that define the contemporary era of formal predictive models.

Framework 1. Cognition as Prediction (Prehistory[1])

As the first and therefore oldest period of prediction, cognition qua prediction is also the framework that is marked by the most uncertainty. Most knowledge from this time period is inferred from the fossil records, limited hominid artifacts, and the comparative study of other primates. More recently the field of neuroscience has found evidence and argued for the idea that prediction may be one of the fundamental functions of the brain (e.g., Bubic, Cramon, & Schuboltz, 2010). Operating under theories with esoteric-sounding names, such as "free energy principle," "predictive coding" and the "Bayesian brain hypothesis," these neuroscientists subscribe to the general notion that:

> "[A]t each level of a cognitive process, the brain generates models, or beliefs, about what information it should be receiving from the level below it. These beliefs get translated into predictions about what should be experienced in a given situation, providing the best explanation of what's out there so that the experience will make sense." (Cepelewicz, 2018)

These theories have given rise to a productive area of research, with eminent neuroscientist Karl Friston's (2010) treatise on the free energy theory receiving over 2,500 citations.

From my perspective, part of the reason for the popularity of these theories is that they tell a compelling evolutionary story. This family of theories has a distinctive functionalist and Darwinist intuition behind it. The gist of the evolutionary rationale behind this group of theories goes as such: A quarter of human caloric intake goes towards feeding our brains (Gibbins, 2016).

---

[1] The date ranges for all three of the frameworks should be considered rough estimates. Additionally, there is never a hard cutoff between rising and falling paradigms of inquiry. Rather, these dates merely indicate periods of relative onto-epistemological dominance. Many people still subscribe to cultural and religious theories of prediction, and I do not intend to make value judgements regarding any of the paradigms being discussed.

Brains must be exceptionally evolutionary adaptive. These caloric requirements would otherwise be untenable. Correspondingly, we can see that minimizing uncertainty through the generation of predictions would also be evolutionary adaptive. By making predictions concerning the future state of the world, the brain allows animals to take actions in a manner that optimizes the chance of survival and reproduction over the long-term, rather than "solely react[ing] to events occurring around us" (Bubic, Cramon, & Schuboltz, 2010). Unlike the similar behavior of "explanation," prediction does not require a conscious understanding *why something* occurs. Rather, prediction merely requires the integration of information through any means possible with the purpose of anticipating what will happen in the future (cf. Yarkoni & Westfall, 2018 for a discussion of the tension between prediction and explanation). As such, it seems feasible that although most animals may not engage in rigorous explanative processes, they most likely do engage in prediction.

As a self-proclaimed "unified brain theory" (Friston, 2010), researchers have found evidence for this hypothesis in a variety of domains. Studies of cue integration have revealed some of the strongest evidence for the Bayesian brain hypothesis. For example, Kira, Yang and Shadlen (2015) trained monkeys to associate series of cues with reward probabilities associated with a response to the monkey's right or left. Kira et al. found that rhesus monkeys were able to integrate probabilistic cues in a near optimal fashion during a decision-making task. They found that, even when the probabilities associated with each cue were manipulated over time, the monkeys were able to update their beliefs and maximize the reward obtained.

Conversely, Friston notes that "if the brain is an inference machine, an organ of statistics, then when it goes wrong, it'll make the same sorts of mistakes a statistician will make" (Cepelewicz, 2018). To prove this point, researchers often point to a variety of perceptual illusions. Most illusions can be understood as the result of pre-existing perceptual biases coming into contact with ecologically invalid stimuli. For example, consider the McGurk effect. McGurk and MacDonald (1976) exposed participants to videos, showing an adult repeatedly uttering the syllable [ba]. Unbeknownst to the participants, the videos had been dubbed with the sound of the adult saying [ga], but the adult’s mouth movements remained unchanged. In a perceptually unbiased world, the participants should have seen [ba], but heard [ga]. Instead, the participants watching the video reported hearing [ba]. This effect has been replicated in many contexts (Tiippana, 2014).

The Bayesian brain framework deftly explains this phenomenon. Throughout their lives, English speakers are exposed to the lip-movements of someone saying [ba] thousands of times. The correlation between these lip-movements and the sound [ba] is nearly perfect. As such, the likelihood of auditorily perceiving [ba] in the presence of corresponding visuals is incredibly high. The McGurk effect arises because our prior predictions regarding the sound we should that should emanate from the video overrides auditory stimuli the participants are truly experiencing.

Friston exists somewhere at the nexus of post-positivism, and functionalism. As neuroscientists by training, the methods of Friston and colleagues are scientific in nature. He is a post-positivist in that he makes use a of a variety of scientific methods to understand the brain. Furthermore, the very nature of the Bayesian brain hypothesis is post-positivist. The Bayesian brain hypothesis assumes that the brain is always taking in new information and updating beliefs (Friston et al., 2015). In that the brain is always taking in new information, updating models, and refining predictions, Friston fundamentally understands the brain as a post-positivist actor.

Taking the Bayesian brain hypothesis to be true, we can see posit that, even before the development of speech, culture, or other schemes of advanced social organization, the brain was making inferences and predictions about the world around it. This hypothesis is supported by research into human (e.g., Friston, 2010) and primate cognition (e.g., Kira et al., 2015). Nevertheless, individualized cognition is not the most efficient way of modeling one's environment, as each individual is only able to predict the outcomes of circumstances they are already at least partially acquainted with. As will be discussed in the next section, the development of language allowed for the sharing and refinement of these cognitively-derived predictions through social interaction.

Framework 2: Informal Belief-Based Frameworks (35,000 BCE – 1500 CE)

Approximately 200,000 years ago, *homo sapiens* developed the ability to communicate using generative language (Pagel, 2017). Soon after the development of language, we see what Michael Corballis (1992) terms "cultural explosion." This cultural explosion is estimated to have begun 35,000 years ago (White, 1985), marking the beginning of what I term the "era of informal socio-religious belief-based frameworks" This era is marked by the dominance of informal belief-based models of prediction. In a departure from the previous era, cognition was no longer individual. Predictions could be shared and refined through dialog. Now, predictions could stem, not only from personal experience, but also from shared socio-cultural-religious experiences.

As mentioned in the introduction, the manner in which doctors diagnose patients is often considered the prototypical example of an informal predictive framework. Most informal predictive models do not require an advanced degree to participate in. For example, consider the driving down a busy street. The only reason that streets function is because drivers and pedestrians have internalized predictions about what every other participant will do. Drivers would not be able to navigate through stop lights if they had expectations that participants on the side of the road would randomly step in front of their car. These predictive frameworks rely on the human ability to preemptively communicate intentions and guidelines. Without language, drivers would have to learn how to drive through trial and error. Needless to say, this would not be a very productive way to learn how to drive.

Although informal predictions are common throughout everyday experience, religion is a particularly strong example of a rigorous, yet still fundamentally informal realm of prediction. Religious predictions are heavily researched and defended, yet seldomly rely on formal mathematical models. As such, I will focus on prediction as it is instantiated in the Bible, not because of any value judgements or belief that Biblical prophecies are inherently any more truthful than those found in other religions, but merely because I happen to be more intimately acquainted this context. In respect to the chosen timeframe of 35,000 BC to 1500 CE, I want to clarify that I do not intend to argue that socio-religious models of prediction are obsolete or incorrect. Rather, my argument revolves around the general trends of "epistemological soft power." I placed the end-date of this era at 1500 CE, not because religion lost its relevance, but purely because alternate methods of prediction began gaining ground in the public sphere.

In particular, I think the Old Testament is a prime example of this era, as it is interwoven with predictions, prophecies, and proverbs. Taking the 46-book configuration used by Catholics, the Old Testament is often split into The Pentateuch, Historical Books, Wisdom Writing, and the Prophets (Boadt, 1984). The Wisdom Writings and Prophets sections are particularly relevant to the theme of prognostication; the titles alone reveal the thread of prognostication that runs through them. The Wisdom Writings are filled with aphorisms. It even includes a book of "Proverbs," including all manners of guidance ranging from meta-advice:

6. For the Lord gives wisdom; from his mouth come knowledge and understanding.
7. He holds success in store for the upright, he is a shield to those whose walk is blameless,
8. for he guards the course of the just and protects the way of his faithful ones.

Proverbs 2:6–8; New International Version

To concretely actionable suggestions, such as:

27. Do not withhold good from those to whom it is due, when it is in your power to act.
28. Do not say to your neighbor, "Come back tomorrow and I'll give it to you"—
when you already have it with you.

Proverbs 3:27–28, New International Version

As signaled by the name of the book and the content therein, the overarching purpose of Proverbs is to provide advice. And advice, independent of its nature, is based off of predictions concerning what actions will lead to favored outcomes. Unlike later predictive frameworks, however, these aphorisms are not based off of data or formal mathematics. Rather the proverbs are given by God, which (at least for believers) lends them a certain level of unassailable truth that human-derived knowledge can never achieve.

Whereas Proverbs is filled implicit predictions (i.e., "follow God and good things will happen"), the sixteen prophets of the Old Testament also gave explicit predictions about the coming nature of the world. The prophecies regarding the coming of the messiah were particularly influential. In fact, these prophecies functioned as some of the foremost indicators of the legitimacy of Jesus. Researchers have estimated that Jesus fulfilled 300 prophecies during his life (Cru, 2018).

Fundamentally, the nature of truth is what separates the paradigm of Christian theology from other theories of prediction. The dominant ontology of Christianity is marked by realism. In 2005, before becoming Pope, Cardinal Joseph Ratzinger declared in his homily that "[w]e are moving toward a dictatorship of relativism which does not recognize anything as for certain and which has as its highest goal one's own ego and one's own desires" (quoted in Wilker, 2013, n.p.). The Catholic Church's critique of relativism stems from its view that relativism will ultimately lead to conflict. In 2013, Pope Francis echoed his predecessor, saying, "[b]ut there is no true peace without truth! There cannot be true peace if everyone is his own criterion, if everyone can always claim exclusively his own rights, without at the same time caring for the good of others…" (Naab, 2013, n.p.). The Pope and by extension the Catholic Church is arguing that they know the truth delivered to them by God and through exegesis, and that there will only ever be peace if the rest of the world joins in this understanding.[2]

Whereas the other frameworks discussed in this paper derive knowledge solely through observation and experimentation, the epistemology of religious frameworks expands upon this narrow conception of what is true. Christian theology accepts that some facts will be discovered through empirical means, but this knowledge is still ultimately dependent on the will of God. This framework also understands that some knowledge that cannot be gleaned through empirical means. This knowledge must be delivered through alternative methods, such as rigorous exegesis of the guiding texts or holy intervention.

This is not to say that holy intervention and the exegesis are requisite aspects of an informal predictive framework. Christianity is merely one example of an informal social framework for prediction, and as discussed in the introduction to this section, informal models of prediction guide our daily lives. Informal models allow us to do everything from driving to grocery shopping. But despite the prevalence of informal models, formal models have been afforded an unmatched level of epistemic legitimacy (Abbott, 1988). This legitimacy has been

[2] Pope Francis has reportedly softened his stance on relativism throughout his tenure, as evidenced by "hardline" reaction to his "apostolic letter on marriage and family, *Amoris Laetitia*" (Mickens, 2016, n.p.).

gained at the highest levels of political and economic decision-making, and the beginning of this rise begins with the work of mechanist philosophers in Europe during the 1500s.

Framework 3. Rise of Formal Predictive Models (1500 CE – Present)

"We are as gods and might as well get good at it."

– Stewart Brand, The Whole Earth Catalog, 1968

Between 1500 and 1800 CE, there occurred substantial development in both the paradigms and tools of prediction. In this interstitial step, work by philosophers and mathematicians, such as Galileo, Descartes, and Laplace, created the necessary preconditions for the development of formal predictive models. Work by mechanists and early empiricists allowed for the gradual shift in the prioritization of empirically-derived rather than culturally-derived knowledge (Petryzak, 2010). These philosophical developments paved the way for empirically-derived predictions.

As with many developments in quantitative methods during this era, the first predictive models were developed in the domain of astronomy. Carl Friedrich Gauss first uncovered the method of least squares in 1795, while attempting to predict the motion of comets around the night's sky (Kopf, 2015). The method of least squares allows for the fitting of a line that minimizes squared error across the dataset being modeled. Gauss's method allows for the prediction of a dependent variable given one or more predictor variables and continues to underlie contemporary linear regression models. Interestingly, Gauss did not consider his method to be particularly innovative and had assumed it to be previously discovered. As a result, Gauss did not publish his findings. In 1805, Adrien Legendare also discovered the method of least squares while studying the movement of comics. Fortunately, Legendre realized the importance of this discovery and published his findings (Kopf, 2015; Denis, 2000).

One might expect that upon the publication of Legendre's results that linear regression would have been an obvious application of the method of least squares and would have been adopted by all manners of scientific research. This was not the case. Little progress was made between 1805 until the late 1800s in the realm of linear regression. It was not until Francis Galton published his work describing the first correlation coefficient in 1888 that rigorous work began in the expansion of linear regression methods (Stanton, 2001). Although Galton put forth the idea of the correlation coefficient, it was his student Karl Pearson who "subsequently made precise many of Galton's ideas" (Blyth, 1994, p. 394; cf. "Professor Karl Pearson," 1936 for a description of the mutual admiration between Galton and Pearson).

Karl Pearson was a prolific statistician, scientist, and philosopher of science.[3] As Floridi (2012) notes, in Pearson's book, *The Grammar of Science* (1892), Pearson advocated for a Kantian-influenced ontology. Both Kant and Pearson were influenced by a combination of rationalist and empiricist views (Barlas & Carpenter, 1990). Like Kant, Pearson subscribed to the idea that there are fundamental truths stemming from an underlying reality. However, unlike Kant, Pearson saw science as framework for understanding reality, rather than objective reality in and of itself. Or as Levine (1996) writes, under Pearson's paradigm "the scientist discovers nothing: he is an inventor" (n.p.)

Under this paradigm, scientific inquiry is needed to minimize uncertainty that results from perception. Or, as Norton (1978) makes clear, Pearson saw that "[a]ny connection, through

---

3 And unfortunately, an avowed racist and eugenicist (Norton, 1978). His prejudice was also shared with his advisor, Francis Galton. In fact, Pearson held the inaugural "Galton professorship of Eugenics" ("Professor Karl Pearson," 1936).

experience, between the self and the real world was therefore highly tenuous, and the only goal for science that made sense was an instrumental one" (p. 14). Looking at Pearson's methodological contributions, this ontological stance makes considerable sense. Pearson's work dealt with the construction of more advanced scientific instruments. These instruments aimed to minimize uncertainty. Yet, Pearson knew that his statistical innovations would never eliminate all uncertainty. For this reason, we also see that Pearson flirted with the ideas of logical positivism. In particular, the idea that certain laws of the universe could be summarized through mathematical means (Norton, 1978).

Pearson certainly appreciated the power of linear regression to minimize uncertainty and make predictions in a variety of domains. However, it is not clear that he would assign them the same level of epistemic legitimacy afforded to them today by scientists and researchers. Although the family of models within the linear regression are used often without much thought within social science and public policy research, Pearson, himself, advocated against the use of linear models within these spheres. In 1894, Pearson noted that, "[i]n the case of biological, sociological, and economic measurements there is ... a marked deviation from the normal shape" (quoted in Blyth, 1994, p. 394). Seeing as non-normality reduces the accuracy of linear regression methods, Pearson is implicitly arguing against the wanton application of linear regression models to social science.

Despite Pearson's admonitions, was quickly adopted by social scientists. By 1915, predictive models had been applied to novel problems in anthropology, sociology, and psychology (Camic & Xie, 1994). Thanks to early work by mathematicians – such as Blaise Pascal, Charles Babbage, Ada Lovelace, and numerous others – infused with unprecedented research and development activity during World War II, the first commercial computers arrived in the 1950s (Mannell, 2009). The proliferation of computational access only hastened the application of predictive models. Since the latter half of the 1960s, social scientists have had access to a wide array of predictive modelling and statistical software (Uprichard, Burrows, & Byrne, 2008).

There has been a concurrent and dramatic increase in the diversity of predictive algorithms. Over the last 60 years, several entire classes of predictive models have been developed and refined. Neural networks first appeared in 1958 (Rosenblatt, 1958). Logistic regression was also first put forth in 1958 (Cox, 1958). Due to an ever-increasing availability of computational hardware, we see a period of further refinement of these techniques during the last two decades of the twentieth century. Entire new classes of predictive algorithms were developed. Random forest models (Ho, 1995) and more advanced neural network algorithms (LeCun, Bottou, Benio, & Haffner, 1998) were both developed in the 1990s. Despite their distinctive names, these algorithms can all be applied to problems of prediction with varying levels of accuracy (Caruana & Niculescu-Mizil, 2006).

Of these families of predictive algorithms, neural networks have received the most popular press within the last decade. Neural networks are a form of formal predictive algorithm that are most often trained using previously collected data labelled with the correct results. Neural networks learn by making predictions. When neural networks make the wrong prediction, the network pinpoints the neurons that led it to make the faulty prediction. The weights associated with these offending neurons are adjusted so that the error between the prediction and true result diminishes.

Created with varying levels of biological plausibility (Bengio, et al., 2015), neural networks at the very least superficially resemble the structure of brains. If we take a Fristonian view of the brain, neural networks also approximate the purpose of the brain. Neural networks and brains both serve as predictive models of the world around them. The brain just happens to be several orders of magnitude more complex. For example, Alpha Go Zero is a neural network algorithm capable of beating grandmaster players of the Chinese board game “go.” Go is significantly more complex than chess. Yet, Alpha Go Zero is composed of only 20,000 neurons (Silver et al., 2017). For comparison, the human brain contains roughly 100 billion neurons (Azevedo et al., 2009).

With this comparison in mind, it is clear why computer scientists are inspired by the progress made by relatively simple neural network models. Yann LeCun, originator of a particularly powerful neural network architecture known as convolutional neural networks, is just one example of several highly cited scientists working in this area. Yet, his h-index is measured at a substantial 107 and his most cited work has garnered 15,000 citations. These citation metrics are just one of many indicators of the “hype” surrounding advances in algorithms for prediction.

Despite the hype associated with recent advances in neural networks and machine learning, most predictive models even those classified as “machine learning” algorithms are fundamentally reducible to regression. Recently, Cheng, Khomtchouk, Matloff and Mohanty (2018) informally proved that neural networks are essentially less interpretable versions of polynomial regression. After a century of technical advancement, machine learning is little more than a nonlinear extension of Karl Pearson’s work.

Framework 4: Formal and Informal Models on the Same Plane (Future?)

This is not to say that I am anti-machine learning or anti-statistics. I am actually quite enamored with the possibilities offered by formal predictive models. Nevertheless, we must be wary of the impact of our tools on our thinking process. In general, statistical models and (post)-positivist paradigms leads us to focus on concepts that can be measured. This is an inherent weakness of formal predictive models, yet one that often goes unmentioned. Furthermore, it is not just the use of statistics that shapes our thinking, but also small choices, such as the use of specific statistical programs that can alter the course of our research. As Uprichard et al. (2008) show with the case of the adoption of SPSS in the sociology, our choice of tools fundamentally alter our approach to science and knowledge construction.

Despite the progress of alternative paradigms since the publication of *The Paradigm Dialog* (Guba, 1990), rarely do quantitative researchers feel the need to defend their research paradigm in the same way that qualitative researchers must (Schwandt, 1990). Following within the footsteps of Guba's *The Paradigm Dialog*, I envision a world where methodological and epistemic assumptions are checked in much the same ways that normality is checked when running an ANOVA. In order for such a world to be achieved, the benefits of informal models need to be explored. Additionally, the benefits of these models must be put on the same epistemic playing field as formal models. Researchers need not give up their post-positivist paradigm in order to embrace the benefits of qualitative research. As Reimer (1996) argues, "[q]ualitative research is a child of the post-positivist world. It is a clear reflection of the mind-set underlying post-positivist beliefs. The most salient of these beliefs for us as researchers is that there are no pure data" (p. 123). To extend this concept, I would add that to be truly post-positivist is to embrace both quantitative (formal) and qualitative (informal) methods of prediction.

One interesting method in which qualitative and quantitative work could be combined is through the use of ensemble models. Ensemble models are predictive models that combine input from two or more distinct models into a singular prediction (Makhtar, 2012). Practically, an ensemble model might take predictions on the same dataset from a linear regression algorithm, a neural network, and a random forest model, weigh these individual predictions against one-another, and ultimately decide upon a final prediction. There is no reason why human input could not be factored into these models as an additional source of information. Given the theoretical and practical concerns surrounding the issue of ethics in computational models (cf. Moor, 2006), Although there is substantial interest in this idea of combining human and artificial predictions (e.g., Amershi, Cakmak, Knox, & Kulesza, 2014; Kamar, Hacker, & Horvitz, 2012), this interest is a relatively small portion of the overall literature on machine learning techniques. For this reason, I fear that this research on the humanization of artificial intelligence may be a passing fad.

Conclusion

"Outcomes are predicted. Trends are forecasted.

Spoilers are included. Major plots are foreshadowed.

In other words, we want to know what's going to happen next."

– *Medium* Email Newsletter

Personal Communication, December 3, 2018

In this paper, I have provided vignettes, showing three distinct modes of prediction: cognitive, belief-based, and formal (statistical). As evidenced by the breadth of frameworks showcased in this paper, predictive models are inherent features of our everyday lives. The majority of models in use today remain informal. Nearly all beliefs produce and are produced by a series of predictions. Simple actions such as driving, shopping at the grocery store, and even reading entail predictions. These beliefs have been communicated to us primarily through socio-religious or perceptual means, not formal mathematical models. Despite the prevalence of informal models in everyday experience, formal models have gained epistemic dominance within many research and academic domains (Abbott, 1988; Guba 1990).

My understanding of prediction has been heavily influenced by the work of Paul Meehl (the formal-informal dichotomy) and Karl Friston (cognition as prediction). Building on these frameworks for understanding prediction, this paper argues that both belief-based and formal models have been boot-strapped upon the original predictive engine – the brain. I have also attempted to afford both formal and informal models the same epistemic legitimacy throughout this paper. These vignettes also serve to complicate the formal-informal model dichotomy, showing that the formal-informal spectrum is orthogonal to the paradigm of inquiry. In other words, both formal and informal models can and are used by post-positivists, critical theorists, religious-based philosophers, functionalists, and constructivists alike. In the fourth vignette, I offered a brief, personal critique of an overreliance on formal predictive models in the post-positivist social sciences. I offer a possible solution to this critique through the combination of informal and formal models in ensemble models. Ensemble models present merely one manner in which formal models could be augmented, enhanced, and made more human.

Altogether, the history of prediction – both in terms of its study and the ways it is enacted – reveals a complex relationship between humans, knowledge, and the future. The vignettes merely served to reveal general trends in this relationship. Clearly, there is need for further research in the trends of prediction, particularly in non-western and prehistorical cultures. From a Fristonian point-of-view, one might say that "to predict is to be human and to be human is to predict." Thus, to understand the human experience, we must understand the frameworks, tools, and cognitive processes underlying prediction.

References

Abbott, A. (1988). Transcending general linear reality. *Sociological Theory*, *6*(2), 169. https://doi.org/10.2307/202114

Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*, *35*(4), 105–120.

Austen, I. (2011, July 25). Early media prophet is now getting his due. *The New York Times*. Retrieved from https://www.nytimes.com/2011/07/26/books/marshall-mcluhan-media-theorist-is-celebrated.html

Azevedo, F. A. C., Carvalho, L. R. B., Grinberg, L. T., Farfel, J. M., Ferretti, R. E. L., Leite, R. E. P., … Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *The Journal of Comparative Neurology*, *513*(5), 532–541. https://doi.org/10.1002/cne.21974

Barlas, Y., & Carpenter, S. (1990). Philosophical roots of model validation: Two paradigms. *System Dynamics Review*, *6*(2), 148–166.

Bengio, Y., Lee, D.-H., Bornschein, J., Mesnard, T., & Lin, Z. (2015). Towards biologically plausible deep learning. *ArXiv Preprint ArXiv:1502.04156*.

Blyth, S. (1994). Karl Pearson and the correlation curve. *International Statistical Review/Revue Internationale de Statistique*, 393–403.

Boadt, L. (1984). *Reading the Old Testament: an introduction*. New York, N.Y: Paulist Press.

Brand, S. (n.d.). We are as gods. Retrieved from http://www.wholeearth.com/issue/1010/article/195/we.are.as.gods

Bubic, A., Cramon, D. Y., & Schuboltz, R. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2010.00025

Camic, C., & Xie, Y. (1994). The Statistical Turn in American Social Science: Columbia University, 1890 to 1915. *American Sociological Review*, *59*(5), 773. https://doi.org/10.2307/2096447

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning* (pp. 161–168). ACM.

Cepelewicz, J. (2018, July 10). To make sense of the present, brains may predict the future. *Quanta Magazine*. Retrieved from https://www.quantamagazine.org/to-make-sense-of-the-present-brains-may-predict-the-future-20180710/#

Cheng, X., Khomtchouk, B., Matloff, N., & Mohanty, P. (2018). Polynomial regression as an alternative to neural nets. *ArXiv Preprint ArXiv:1806.06850*.

Corballis, M. C. (1992). On the evolution of language and generativity. *Cognition*, *44*(3), 197–226.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, 215–242.

Cru. (2018, January 4). 55 Old Testament prophecies about Jesus. Retrieved from https://www.jesusfilm.org/blog-and-stories/old-testament-prophecies.html

Denis, D. (2000). *The origins of correlation and regression: Francis Galton or Auguste Bravais and the error theorists?* Retrieved from https://www.york.ac.uk/depts/maths/histstat/bravais.htm

Floridi, L. (2012). Degenerate epistemology. *Philosophy & Technology*, *25*(1), 1–3. https://doi.org/10.1007/s13347-012-0067-6

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, *6*(4), 187–214. https://doi.org/10.1080/17588928.2015.1020053

Gibbons, A. (2016). Humans are the highest energy apes, making us smarter—but also fatter. *Science*. https://doi.org/10.1126/science.aaf5692

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*(1), 19–30. https://doi.org/10.1037//1040-3590.12.1.19

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, *2*(2), 293.

Guba, E. G. (1990). The paradigm dialog. In *Alternative Paradigms Conference, Mar, 1989, Indiana U, School of Education, San Francisco, CA, US*. Sage Publications, Inc.

Ho, T. K. (1995). Random decision forests. In International Association for Pattern Recognition (Ed.), *Proceedings of the Third International Conference on Document Analysis and Recognition: August 14 - 16, 1995, Montréal, Canada*. Los Alamitos, Calif.: IEEE Computer Society Press.

Howarth, R. J. (2001). A history of regression and related model-fitting in the earth sciences (1636?-2000). *Natural Resources Research*, *10*(4), 241–286.

Kamar, E., Hacker, S., & Horvitz, E. (2012). Combining human and machine intelligence in large-scale crowdsourcing. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1* (pp. 467–474). International Foundation for Autonomous Agents and Multiagent Systems.

Kira, S., Yang, T., & Shadlen, M. N. (2015). A neural implementation of Wald's sequential probability ratio test. *Neuron*, *85*(4), 861–873. https://doi.org/10.1016/j.neuron.2015.01.007

Kopf, D. (2015, November 6). The discovery of statistical regression. Retrieved from https://priceonomics.com/the-discovery-of-statistical-regression/

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*, *98*(6), 1060–1072. https://doi.org/10.1037/a0034156

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Levine, G. (1996). Science and citizenship: Karl Pearson and the ethics of epistemology. *Modernism/Modernity*, *3*(3), 137–143. https://doi.org/10.1353/mod.1996.0046

Lewis, E. C., & Mackinney, A. C. (1961). Counselor vs. statistical predictions of job satisfaction in engineering. *Journal of Counseling Psychology*, *8*(3), 224.

Makhtar, M. (2012). *Contributions to ensembles of models for predictive toxicology applications: On the representation, comparison and combination of models in ensembles*. Retrieved from http://hdl.handle.net/10454/5478

Mannell, R. (2019, December 14). A short history of computers and computing. Retrieved from http://clas.mq.edu.au/speech/synthesis/history_computers/index.html

Meehl, P. E. (1954). Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.

Mickens, R. (2016, June 1). Letter from Rome: Is Pope Francis a relativist. *Commonweal Magazine*. Retrieved from https://www.commonwealmagazine.org/letter-rome-81

Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems*, *21*(4), 18–21.

Naab, K. (2013, March 22). Francis follows Benedict in decrying 'tyranny of relativism.' *Zenit*. Retrieved from https://zenit.org/articles/francis-follows-benedict-in-decrying-tyranny-of-relativism/

Norton, B. J. (1978). Karl Pearson and Statistics: The Social Origins of Scientific Innovation. *Social Studies of Science*, *8*(1), 3–34. Retrieved from http://www.jstor.org/stable/284855

Pagel, M. (2017). Q&A: What is human language, when did it evolve and why should we care? *BMC Biology*, *15*(1). https://doi.org/10.1186/s12915-017-0405-3

Pearson, K. (1892). *The grammar of science*. Walter Scott.

Petryzak, N. (2010). Tabula rasa – Its origins and implications. In L. T. Benjamin Jr (Ed.), *A history of psychology: Original sources and contemporary research*. (3rd ed.). Mcgraw-Hill Book Company.

Professor Karl Pearson, F.R.S., 1857-1936. (1936). *Journal of the Royal Statistical Society*, *99*(2), 424–428. Retrieved from http://www.jstor.org/stable/2980604

Rao, V. (2018, November 21). The age of early divinity. Retrieved from https://www.ribbonfarm.com/2018/11/21/the-age-of-early-divinity/

Reimer, B. (1996). Qualitative research and the post-positivist mind. *Bulletin of the Council for Research in Music Education,* (130), 123-126. Retrieved from http://www.jstor.org/stable/40318818

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386.

Schwandt, T. R. (1990). Paths to inquiry in the social disciplines: Scientific, constructivist, and critical theory methodologies. *The Paradigm Dialog*, 258–276.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., … Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, *550*(7676), 354–359. https://doi.org/10.1038/nature24270

Sniderman, A. D., D'Agostino Sr, R. B., & Pencina, M. J. (2015). The role of physicians in the era of predictive analytics. *JAMA*, *314*(1), 25. https://doi.org/10.1001/jama.2015.6177

Stanton, J. M. (2001). Galton, Pearson, and the peas: A brief history of linear regression for statistics Instructors. *Journal of Statistics Education*, *9*(3). https://doi.org/10.1080/10691898.2001.11910537

Tiippana, K. (2014). What is the McGurk effect? *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00725

Truesdell, A. B., & Bath, J. A. (1957). Clinical and actuarial predictions of academic survival and attrition. *Journal of Counseling Psychology*, *4*(1), 50.

Uprichard, E., Burrows, R., & Byrne, D. (2008). SPSS as an 'inscription device': From causality to description? *The Sociological Review*, *56*(4), 606–622. https://doi.org/10.1111/j.1467-954X.2008.00807.x

van Otterlo, M. (2014). Automated experimentation in Walden 3.0: The next step in profiling, predicting, control and surveillance. *Surveillance & Society*, *12*(2), 255.

Westen, D., & Weinberger, J. (2004). When clinical description becomes statistical prediction. *American Psychologist*, *59*(7), 595.

White, R. (1985). Thoughts on social relationships and language in hominid evolution. *Journal of Social and Personal Relationships*, *2*(1), 95–115.

Wiggins, N., & Kolen, E. S. (1971). Man versus model of man revisited: The forecasting of graduate school success. *Journal of Personality and Social Psychology*, *19*(1), 100–106. https://doi.org/10.1037/h0031147

Wilker, B. (2013, February 25). Benedict vs. the dictatorship of relativism. *National Catholic Register*.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Appendix A: Pivotal Events in the History of Prediction

Framework 1. Cognition as Prediction (Prehistory)

- Evolutionary Divergence Between Humans and Closest Ancestors
- Development of Language
- 2010s – Karl Friston puts forth the free energy principle, a unifying theory of the brain emphasizing prediction.

Framework 2. Informal Belief-Based Frameworks (35,000 BCE – 1500 CE)

- 35,000 BCE – Artifacts indicate presence of "cultural explosion." This process results in a multitude of powerful social belief systems, both religious and non-religious.
- 6 BCE until 36 CE – Life of Jesus Christ, Founding of the Catholic Church
- 1500s – Mechanist philosophy arrives in the West.
- 2013 – Popes Benedict and Francis speak out against relativist paradigms.

Framework 3. Rise of Formal Predictive Models (1500 CE – Present)

- 1500 until 1650 – Mechanist philosophers, such as Galileo and Descartes, lay the epistemological groundwork for the coming empiricist revolution.
- 1777 until 1855 – Life of Carl Friedrich Gauss, who developed the method of least squares in order to describe the movements of the solar system. This method continues to underlie modern regression algorithms.
- 1815 until 1852 – Life of Ada Lovelace, who published the first formulation of an algorithm designed for a computer.
- 1819 – Charles Babbage begins constructing the difference engine, the earliest automatic calculator and predecessor to the modern computer.
- 1857 until 1936 – Life of Karl Pearson, developer of the chi-squared test, the Pearson correlation coefficient, and p-values.
- 1957 – First implementation of Frank Rosenblatt's perceptron, the earliest neural network architecture.
- 1980 until 1998 – Advances by Yann LeCun and colleagues lead to the development of powerful convolutional neural network models, which were modeled off of the feline visual system.
- 2000s until Present – Development of prediction algorithms continues to advance.

Appendix B: Resources

Category	**Example**
Published Empirical article (the last 5 years)	Kira et al. (2015)
Published Empirical article (before 10 years ago)	McGurk and MacDonald (1976)
Published Conceptual article (before 10 years ago)	Friston (2012)
Unpublished thesis or dissertation	Makhtar (2012)
Pop psychology resource	Cepelewicz (2018)
Newspaper or magazine article within the last 10 years	Austen (2011)
A biography or autobiography (in any form: book, article, media)	Norton (1978)
An interview transcript or part of a transcript	Wilker (2013)
An obituary	"Professor Karl Pearson" (1936)
A letter	Mickens (2016)
Social Sciences Citation Index	Karl Friston and Yann LeCun