EXPLORING HINDSIGHT BIAS USING COMPUTATIONAL MODELING AND MACHINE LEARNING CLASSIFICATION IN DIFFERING DOMAINS

by

CAROLINE HIXON

(Under the Direction of Adam Goodie)

ABSTRACT

Hindsight bias occurs when people overestimate what was known before an event once the outcome is known. This work describes an application of a specific cognitive model, Reconstruction After Feedback with Take the Best (RAFT), as a computational model. It takes the form of a supervised classification model utilizing K-Nearest Neighbors, Decision Tree, and Random Forest machine learning techniques, and is implemented in two settings, novel COVID-19 and everyday nutrition. Testing reveals outcomes are dependent on the domain. Nutrition is found to be associated with greater hindsight bias than COVID-19, and COVID-19 deaths to be associated with greater hindsight bias than cases. The results are discussed with regard to their potential applications for hindsight bias literature and the field of machine learning. The results are discussed with regard to their potential applications for hindsight bias literature and the field of machine learning.

INDEX WORDS: Hindsight bias, machine learning, classification, supervised learning, RAFT, COVID-19, nutrition

EXPLORING HINDSIGHT BIAS USING COMPUTATIONAL MODELING AND MACHINE LEARNING CLASSIFICATION IN DIFFERING DOMAINS

by

CAROLINE HIXON

B.S., Tulane University, 2019

A Thesis Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the Degree.

MASTER OF SCIENCE

ATHENS, GEORGIA

2021

EXPLORING HINDSIGHT BIAS USING COMPUTATIONAL MODELING AND MACHINE LEARNING CLASSIFICATION IN DIFFERING DOMAINS

by

CAROLINE HIXON

Major Professor: Adam Goodie

Committee: Frederick Maier
Kimberly Van Orman

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
December 2021

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Dr. Goodie, for teaching me and guiding me throughout this entire process. I thank him for his continued dedication and patience, especially in giving me feedback on the small things.

I would also like to thank Dr. Maier and Dr. van Orman, my remaining committee members. Their involvement, feedback, and comments were instrumental in creating this thesis, and it would not be the same without them.

Lastly, I am very grateful to my family for their unwavering love and support. Thank you to my parents, for giving me the determination to keep learning.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

It is common to be confident in one's own ability to make and remember decisions accurately. Consider the following scenarios:

Two people are discussing the weather and come to the agreement that the upcoming tropical storm will only mildly impact people with thunderstorms. After the tropical storm, which destroys many homes in the area, the two people then decide that in fact, they both knew this was going to be the worst storm of the year.

In a local election, a person votes for the candidate they prefer while asserting their candidate definitely will win. During the tallying of votes, the other candidate acquires a sizable lead, which causes this person to doubt if their candidate will win. After all the votes have been counted, the person's candidate does manage to win, to which the person exclaims they were always sure their candidate would be the winner.

What do these scenarios have in common? For one, both include events that change the relevant information of the scenario. Next, both have a sense of certainty, before and after the event. And these two cases also include reflection upon a previous assertion. Essentially, both of these scenarios outline hindsight bias.

Hindsight bias, also known as the "knew it all along" phenomenon, occurs when once the outcome of an event is known, people tend to overestimate what was known before the event and the predictability of the event's outcome. This cognitive bias emphasizes the complexity of accurately remembering or reconstructing one's ability to predict an event after the outcome, new information, or a correct answer is revealed.

Hindsight bias has been shown to occur pervasively in all types of decision making or judgment tasks in various domains. Some of these domains include elections, labor disputes, terrorist attacks, betting, accidents and injuries, and medical diagnoses (Berstein et al, 2001, Hawkins & Hastie, 1990, Roses & Vohs, 2012, Meuer et al, 2021). The kind of effects ensuing from hindsight bias offers one reason as to why there is such substantial literature surrounding this phenomenon. For example, hindsight bias has been shown to cause overconfidence in decisions, an inflation in perception of self, or alter decision making ability (Roese & Vohs, 2012, Fischoff & Beyth, 1975, Pezzo, 2011). With these effects in mind and because of its widespread relationship with cognition, hindsight bias has been studied in depth, both in the real world and in controlled laboratory settings. The core examinations focus on understanding the causes and effects of this phenomenon.

While highly studied, there is not a singular agreed upon cognitive model, or a theoretical representation of how the mind works, that fully captures all facets of hindsight bias. Similarly, there is not an agreed upon computational model, which is a model that uses computers to simulate and study complex systems. In attempting to better understand and explain the cognitive process and implications of this phenomenon, this paper implements one of the established cognitive processes found in the literature, Reconstruction After Feedback with Take the Best (RAFT), that focuses on recollection and reconstruction biases associated with hindsight bias. The following sections of this chapter go through the literature on hindsight bias, the origins of hindsight bias, and the motivations and contributions of this paper.

1.1 Hindsight Bias in Literature

In 1975, Baruch Fischhoff was the first to formalize the pervasive cognitive distortion now known as hindsight bias (Fischhoff, 1975) in psychological research. He designed a study that sought to observe how accurately people recorded having knowledge at a previous time. He presented two groups of participants each with different scenarios, one that included foresight with a list of possible outcomes, and one with hindsight that included the "correct" outcome that was supposed to be ignored. Overall, Fischhoff found that those in the hindsight group could not ignore the given outcome and gave higher likelihood estimates to that outcome. Since this first study, hindsight bias has been investigated in multiple domains and across multiple disciplines.

Investigations mainly follow two experimental paradigms where hindsight bias is identified and measured (Bernstein, 2016, Pohl, 2007, Mahdavi et al., 2017). In a *hypothetical-outcome design*, hindsight bias is found when two different groups of people have differences in judgements, similar to Fischhoff's original study design (Fischhoff, 1975). The group given the correct answer is told to answer what they would have if they had not been given the answer. Here hindsight bias is observed when the group with hindsight information gives answers closer to the correct outcome. In a *memory design*, there is only one group of participants and one condition instead of two. Participants are asked a question and offer an original response; then researchers reveal the true outcome, and ask the participants what the original answer was. Here hindsight bias is observed when the post-information answer moves closer to the correct answer as compared to the original.

From the substantial literature since 1975, two principal conclusions have been drawn about hindsight bias. Firstly, that it is robust (Pezzo, 2011, Pohl, 2007). This is due to the fact that hindsight bias is found in various settings, such as with stock markets (Knoll et al., 2017), medical diagnoses (Hugh et al., 2009), election predictions (Meuer et al., 2021), and auditing judgments (Bernstein et al., 2016). Additionally, hindsight bias can occur at any age, within any culture, and with any intelligence level (Bernstein et al., 2011).

Secondly, there is a general agreement in the literature that hindsight bias and its robustness are a result of cognitive processes. Cognitive processes are any of the mental functions associated with accessing, storing, or manipulating knowledge. Beyond this general agreement for the prominence of cognitive processes, there is not a consensus upon a single model for them. Instead, the most pervasive of the theories for the cognitive processes surrounding hindsight bias can be represented in three main categories: memory, reconstruction, and motivation (Bernstein et al., 2016).

One category encompasses the theories having to do with *memory*, specifically how issues with memory translate to hindsight bias. In this category, hindsight bias can result from biased recall, possibly due to automatic, unconscious assimilation of the correct answer (Fischhoff, 1975) or from differing strengths of memories (Hell et al., 1988).

Another category of the cognitive processes behind hindsight bias emphasizes how bias appears when accessing memory fails, instead leading to the reconstruction of the original answer. This *reconstruction* bias is found when the reconstruction of the original response is closer to the correct response than originally reported. Additionally, this category emphasizes how hindsight bias is gradual, meaning the degree of change toward the correct answer shows the extent of one's hindsight bias (Schwarz & Stahlberg, 2003, Roese & Vohs, 2012). Bias in reconstruction can result from the correct answer functioning as an anchor (Fischhoff, 1975, Bernstein et al., 2016, Mahdavi et al., 2017). The anchoring and adjustment theory led to the development of the Selective Activation and Reconstructive Anchoring (SARA) cognitive model. Here, the correct answer alters the pattern of associations for memory traces, meaning the correct answer anchors or guides reconstruction (Bernstein et al., 2016, Pohl et al., 2000). Another possible cause for hindsight bias from reconstruction argues that when memory retrieval is not possible, the original answer is reconstructed with bias because of using foresight and hindsight information. The formal model following from this theory is Reconstruction After Feedback with Take the Best (RAFT) (Bernstein et al., 2016, Hoffrage & Pohl, 2003, Hertwig et al., 2003). Another possibility for hindsight bias in reconstruction is based on metacognition, or thinking about thinking. Here, judgments can affect confidence in original and recalled answers (Müller & Stahlberg, 2007, Sanna & Schwarz, 2007). The "knew it all along" feeling is a reflection of overconfidence in one's ability to make or remember a decision. In contrast with SARA and RAFT, which are considered automatic processes, many of the metacognitive motivations are considered effortful and conscious.

The third category of all cognitive processes causing hindsight bias surrounds the role of *motivation*. Motivational factors, such as aiming to maintain one's self-worth as taken from perceived levels of intelligence and memory capabilities, are thought to be a foundation for this category. Specifically, motivations reflect hindsight bias when moving toward the correct answer occurs in the face of negative outcomes (Blank & Nestler, 2007, Bernstein et al., 2016, Louie, 1999, Ash and Wiley, 2008) or in response to previous bad decisions (Louie, 1999, Roese & Vohs, 2012, Coolin et al., 2015, Ash, 2009).

The three categories of cognitive processes in hindsight bias literature suggest the difficulty in capturing the robustness and complexity of hindsight bias in a single model. Different situations and settings generate the necessity of utilizing different cognitive processes. Instead of one of these categories being the most suitable for every context, it is likely that all of these categories and processes play a role in producing hindsight bias. The next section discusses how the past literature contributes to the present paper.

1.2 Motivations and Contributions

There is an abundance of literature and psychological experimentation investigating the various cognitive processes behind hindsight bias. Yet the literature lacks research into computational models, or models that use computers to simulate and study complex systems, of these cognitive processes. Additionally, the literature does not provide a unified, comprehensive model for understanding, interpreting, or predicting hindsight bias.

This paper presents an application of a combination of the prominent cognitive processes underlying hindsight bias in a way that accounts for two of the three theoretical umbrellas of hindsight bias models. Specifically, this paper's model accounts for distorted memory and reconstruction in the face of uncertainty and different settings. The two settings, utilized in order to analyze how context influences hindsight bias, are the urgent and novel COVID-19 and the common, everyday nutrition.

From this computational model, I aim to further understand the underlying cognitive processes of hindsight bias in a way that can provide a framework for better, cohesive model of hindsight bias as well as providing a distinction of conditions where hindsight bias is present or absent. Beyond contributing to the expansive literature on hindsight bias, this paper hopes to inform how understanding hindsight bias's cognitive processes can translate into identifying and preventing the same biases in machine learning, leading to greater explainabilty for machine learning algorithms.

The following chapter details this paper's model design which utilizes Hoffrage's Reconstruction After Feedback with Take the Best (RAFT) cognitive model.

CHAPTER 2

MODELING HINDSIGHT BIAS

This paper's computational model accounts for two of the three categories of cognitive processes thought to underlie hindsight bias: memory and reconstruction. Since this model is framed as following the memory design of the two experimental paradigms, the model is asked a question, given some information, and asked to remember the original response. Memory issues are introduced in the form of time and uncertainty, where the original and subsequent decisions are influenced by a degree of time between them and the possibility that original memory is unreliable. To account for reconstruction in the face of irretrievable memory, the RAFT model is adopted.

The two categories of cognitive processes chosen for this paper's model are relevant since they offer access to different knowledge. When a memory is recalled, the strength of the memory is considered. With RAFT, a memory is reconstructed with the original and new information that recall does not incorporate.

2.1 RAFT Explained

The model introduced follows the RAFT method, Reconstruction After Feedback with Take the Best, first introduced by Hoffrage (Hoffrage & Pohl, 2003, Bernstein et al., 2016, Blank & Nestler, 2007). In this cognitive model, which follows the memory design experimental paradigm, there are three time steps. First, a task and context information are given at Time 1 (T1), where an original decision is made. Following the decision at T1, new information is introduced at Time 2 (T2). Finally, at Time 3 (T3), the original response must be recalled.

The RAFT method follows from the theory of probabilistic mental models (PMM) (Hertwig et al., 2003, Hoffrage & Pohl, 2003). This theory describes the cognitive process behind decision-making tasks that are between choices. Essentially, this theory formalizes that making a choice based on a quantitative criterion results in a confidence value for that choice. RAFT builds from the PMM theory by utilizing both the confidence value and original choice in reconstruction.

RAFT is utilized in this paper because of the three distinct assumptions it makes. Firstly, if the original response made cannot be retrieved in memory, the original response will be reconstructed by reanalyzing the original problem. This reflects the complexity associated with hindsight bias since simply recalling an old belief is not always possible. Second, the reanalysis that occurs if the original response is irretrievable involves the cues and cue values used in the original choice. Naturally, the original decision influences the reconstructed decision. The third assumption follows that knowledge, in particular uncertain knowledge, is automatically updated by the given feedback. This feedback indirectly influences or changes the knowledge used in the reconstruction process, meaning that while this process enables individuals to improve their inferences over time, it has a by-product: hindsight bias.

2.2 Take the Best Explained

In the decision-making process at Time 1, and later if applicable at Time 3, the RAFT method uses an inferential heuristic to pick which cue has the most predictive power in order to make a decision. This heuristic is named Take the Best and is the heuristic most commonly associated with the PMM framework (Gigerenzer & Todd, 1999, Hoffrage & Pohl, 2003, Hertiwg et al., 2003). This heuristic performs as follows: given cues with their associated validities, an inference is made based on the cue with the highest discrimination value. First, the cue values are searched in memory. If the best cue discriminates, searching is stopped; if the best cue does not, search is continued until the next best is found. Once searching stops, the decision about the cue is chosen; if there is no cue that discriminates, a random choice is made. Finally, the cue validity is used to update the confidence value of the choice, where a random guess is simply 50%.

This heuristic is used because of its simplicity and functionality as a one-reason decision making method, meaning it decides from only one cue, i.e. the best. Due to its simplicity, it has great efficiency, which is important in complex deliberation models. Thus, Take the Best is known as a fast and frugal heuristic from its quick computational power and since it only searches through some of the available information. It has been found to generally provide higher accuracy than other fast and frugal heuristics with one-reason decision making (Gigerenzer & Goldstein, 1999).

2.3 Complete Cognitive Process

The comprehensive cognitive process that serves as a foundation for this paper's computational implementation is summarized in figure 2.1. It starts at Time 1.

Time 1: Original Response

> At T1, there is a certain amount of given available information and a question is asked. This question requires a binary response. Using the information and the Take the Best heuristic, an original response (OR) is decided.

Time 2: Feedback

> Sometime after the original response is made, new information is added to what was originally available. There is access to the combined original and new information at Time 3.

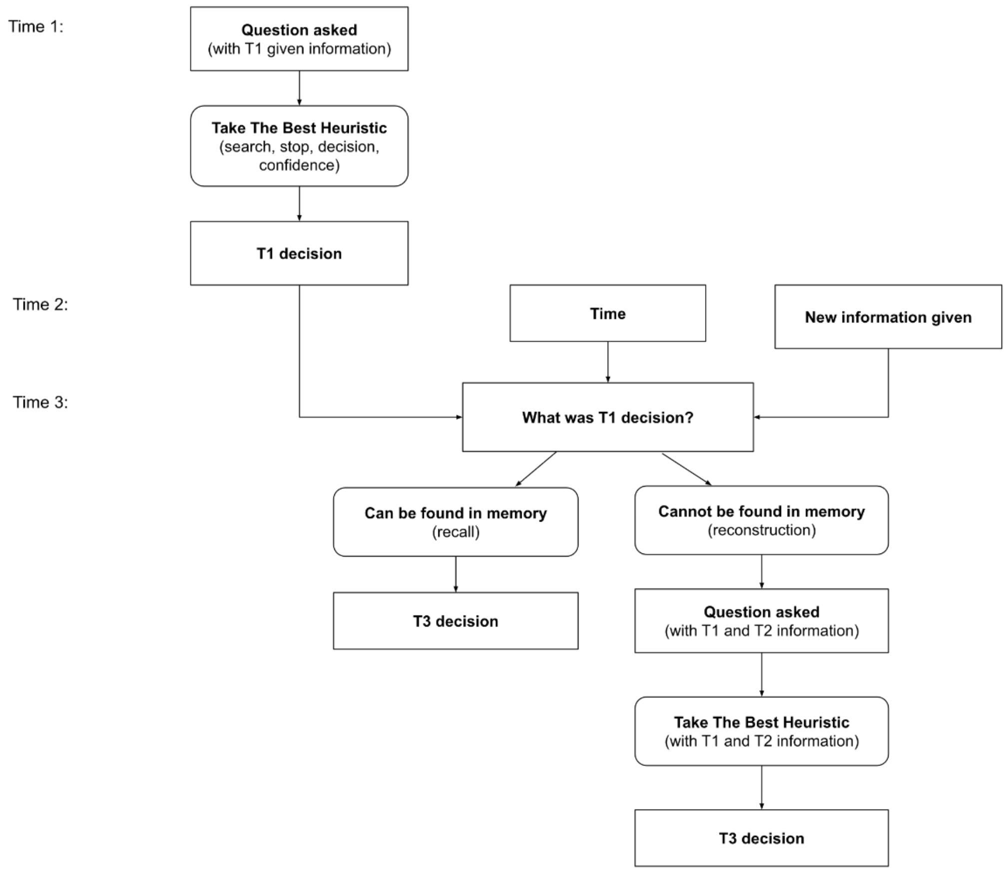

Figure 2.1: Summarization of complete cognitive process following RAFT.

Time 3: Recall or Reconstruction

After the information update, there is a reflection on the original choice. With the time introduced, there will either be a recollection (ROR) of the original choice or a reconstruction (ROR) of the choice if it is inaccessible in memory. If retrievable, the confidence value of the choice is updated with the amount of time between T1 and T3, meaning the model will reflect on the strength of its memory. If the original response is irretrievable in memory, steps from T1 are reconstructed, but now using the combined accessible information. The reconstruction process follows the steps of T1 where the information and TTB heuristic results in a T3 decision.

From the formalization of this cognitive process, hindsight bias is measured by comparing the decisions at T1 and T3 in multiple ways. One way is by comparing the original response to the recalled or reconstructed response (ROR = OR). Hindsight bias is present when the decisions are strengthened or completely changed. For example, the confidence value in the OR is strengthened when it is discovered that the original response was correct. Another way to measure hindsight bias is between the recalled or reconstructed response and the correct response (ROR = CR). Hindsight bias is present when the ROR is the same as or moves towards the CR after learning new information. The formulation for these measurements is shown in figure 2.2.

Finally, it is possible to measure hindsight bias using the OR, CR, and ROR in the following way:

$$\triangle HB = |OR - CR| - |ROR - CR| \tag{2.1}$$

With this equation, if one does not engage in hindsight bias, their original response and their recalled or reconstructed response will be approximately the same, meaning the change in hindsight bias should be close to 0. Alternatively, if one does suffer from hindsight bias, their recalled or reconstructed response will shift towards the correct response, regardless if the original response is the same as the correct response. This means the change in hindsight bias should be positive and larger than the case without hindsight bias.

The next chapter introduces the implementation of this cognitive process as a computational model. Specifically explaining how this cognitive process translates to the machine learning context and offering a description of the two different domains the model is applied to.

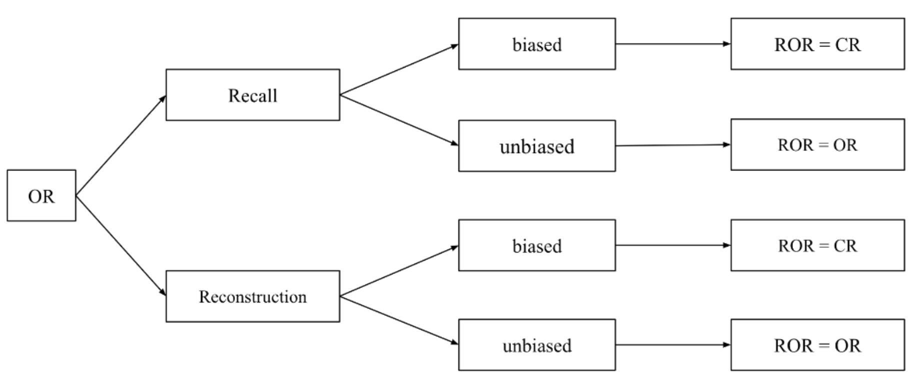

Figure 2.2: Decisions made with hindsight bias after original response (OR). Each branch gives corresponding output, used to formulate hindsight bias equation.

CHAPTER 3
DATA AND METHODOLOGY

3.1 Data Summarization

This project utilizes many different datasets with the aim of assessing hindsight bias in two different domains.

3.1.1 COVID-19

First, hindsight bias is examined in the context of COVID-19. With the increase in cases, deaths, and hospitalizations of the novel Coronavirus, effects have spread beyond the medical community. Indeed, COVID-19 has impacted social, domestic, and economic spheres across cultures and countries (Brennan, 2021). This unique setting provides an abundance of factual and subjective information. Media coverage and public perception are strongly related to the coverage of COVID-19, making it possible here to incorporate subjective information with factual information.

In the context of this project's computational model, the model is given a combination of factual and subjective reports and asked to decide if either cases or deaths will increase or decrease over a given period of time.

There are three sources of data relating to COVID-19 utilized in this paper: the Center for Disease Control (CDC), Consumer Reports of American Experiences Survey (Consumer Reports), and surveys of consumer behavior in the time of COVID-19 from IPSOS (IPSOS). The CDC offers the relevant factual data where the surveys offer subjective data. The CDC provides a tracker dataset that has updated case and death counts daily since January of 2020. The format of the tracker dataset includes attributes for each day for each territory in the United States. The Consumer Reports offer qualitative data in the form of attitudes about various COVID-19 based questions. The IPSOS surveys are also included because they recorded COVID-19 attitudes framed in a different way, providing more data as well as the potential for better representation and generalization. A full description of the COVID-19 sources is given in Appendix A.

3.1.2 Nutrition

The second domain for this paper's model is in the context of an everyday, commonly observed decision concerning healthy food choices. Nutrition data offers factual information about various foods, specifically lacking novelty and social urgency like associated with COVID-19. In this setting, the model is asked to decide if a given food is healthy or not.

The nutrition data comes from the United States Department of Agriculture (USDA) National Nutrient Database for Standard Reference (Haytowitz et al., 2019). This dataset functions as the major source of factual food composition data in the United States which includes 8,617 entries with 45 different attributes for each. The relevant subjective data about healthy food choices comes from the International Food Information Council (2021). This survey from May 2021, offers information on Americans perceptions, beliefs, and behaviors surrounding food. It gives details about health and nutrition as well as eating patterns and purchasing behaviors. A full description of the nutrition datasets is included in Appendix A.

3.2 Preprocessing

This section describes the methods behind cleaning, feature selection, engineering, and scaling of the datasets, and offers a comprehensive view of the final datasets used in this paper. First, the COVID-19 data is described, then the nutrition data.

In order to clean the data, apply preprocessing, build the computational model, and incorporate machine learning models, the Python programming language (Python Software Foundation) is used within the Anaconda environment (Anaconda Software Distribution). Throughout this process, the built in packages of Pandas (MicKinney, 2010), Scikit-Learn (Pedregosa et al., 2011), matplotlib (Hunter, 2007), numpy (Oliphant, 2006), random (Van Rossum, 2020), and datetime (Van Rossum, 2020) were used.

3.2.1 COVID-19

3.2.1.1 Cleaning and Feature Selection

In order to combine the three COVID-19 based datasets, first each dataset needs to be examined and cleaned individually. The process of cleaning entails fixing or removing incorrect, duplicate, or missing data. The CDC tracker dataset, that holds an entry for each day of case and death tabulations, does not require any cleaning since there are not any missing or outlier values. For the Consumer Reports surveys, there are three entries (weeks) missing within the time frame of the tracker dataset. The missing weeks' concern values are fixed by averaging between the previous and following week's scores. The IPSOS data does not have any missing values in the time frame, thus does not require the same averaging for missing scores as the Consumer Report surveys.

The next step in preprocessing is to apply feature selection, or determine which attributes are relevant to the problem. Here, the irrelevant attributes are dropped from the dataset. All attributes are preserved from the CDC tracker dataset. In both survey datasets, many questions are not relevant to the problem, so these two datasets have the most dropped attributes. From the Consumer Reports, only two attributes are selected, concern over the next month and concern of the next 6 months. Similarly, from the IPSOS survey dataset, only one attribute is selected, the personal perceived threat of COVID-19. With cleaning and feature selection complete, the next section describes how the dataset is engineered, from decompositions to transformations and aggregations.

3.2.1.2 Feature Engineering and Feature Scaling

This step in preprocessing requires the most attention. In the tracker dataset, the first action is to transform it so that there is one entry per day that accounts for the entirety of the United States, instead of one entry per day per territory. Here, the data is grouped first by date, then the counts of all remaining attributes (total cases, total deaths, new cases, new deaths) are summed, creating one instance for each date instead of one for each of the 52 jurisdictions.

After each attribute is summed, two new types of attributes are created that reframe the dataset's numerical quantities. One type calculates the difference in counts from the previous day, labeled Case Diffs and Death Diffs. From this attribute, the second attribute of binary output label is assigned for both cases and deaths. If the difference from the previous day is negative, the value is 0 (decrease) and if the difference is zero or positive, the value is 1 (increase).

At this point in the feature engineering, all relevant attributes are combined into one dataset. Since each dataset has the same number of instances, with each being a different day, the datasets were simply merged. The final step in feature engineering is the decomposition of the datetime feature. Five new datetime numerical attributes are created: day, month, year, week day, and week of month.

Feature scaling, or standardizing the numerical quantities of each attribute to the same fixed range, is an important part of preprocessing. Descriptive histograms reveal the attributes ‘newCases’ and ‘newDeaths’ are both tail heavy. This makes sense since the growth in cases and deaths was originally higher during the beginning of the pandemic in 2020. The skewed distributions show that every attribute does not conform to the same scale, making this step of preprocessing necessary. The combined dataset is normalized with Scikit-Learn's preprocessing Normalizer that results in each attribute’s values being in the range from 0 to 1.

3.2.1.3 Overview of complete COVID-19 dataset

After all the preprocessing, the final COVID-19 dataset, now referred to as the COVID-19 dataset, has 17 columns and 541 entries. All attributes are integers except for the full datetime object. The distribution of output labels is shown in figure 3.1 for both cases and deaths, revealing that in both situations the output labels are almost balanced, and that there are more instances of cases increasing, while there are more instances of deaths decreasing. The COVID-19 dataset is not split into a train and test set during preprocessing because the split is dependent on the model parameters, as explained in the next chapter. Instead, at the time of implementation the dataset is split into a train and test set, and then a train and test validation set.

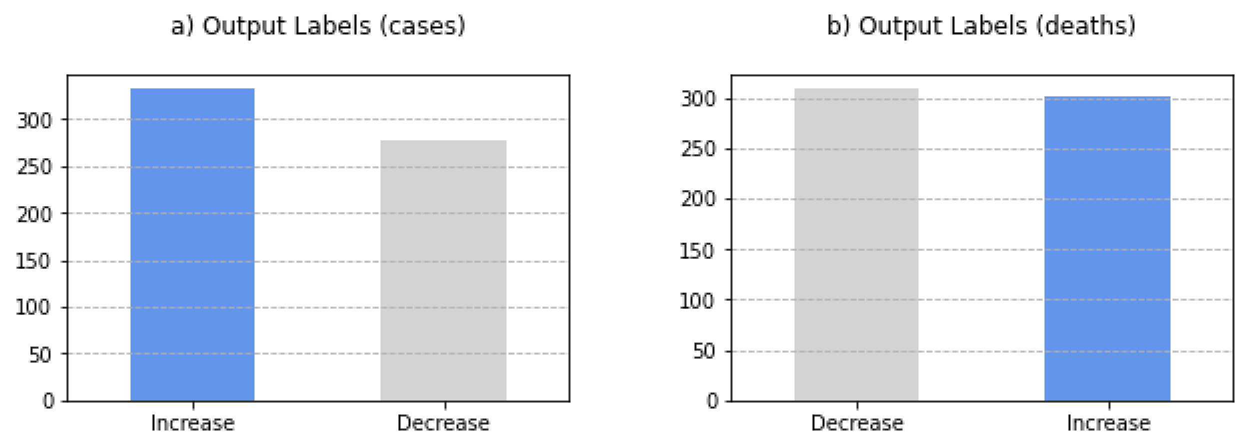

Figure 3.1: Overview of COVID-19 dataset. a) shows more instances of increase for cases, b) shows more instances of decrease for deaths.

3.2.2 Nutrition

In this subsection, the same cleaning and engineering process is described for the nutrition dataset.

3.2.2.1 Cleaning and Feature Selection

In order to join the two nutritional datasets, each needs to be examined and cleaned individually. Within the USDA nutritional dataset, analysis reveals that all quantitative attributes do not have any missing values or outliers, meaning this dataset does not need to be cleaned. The survey dataset reveals the same. The cleaning process is simpler as compared to the COVID-19 process since this dataset does not incorporate datetime attributes.

Concerning feature selection, each dataset retains different attributes. The USDA dataset drops most non-integer attributes, such as 'ShortDescrip' and 'Descrip'. From the survey dataset, two attributes are preserved: what kinds of foods are most likely to cause weight gain and what defines a food as healthy. With cleaning and feature selection complete, the next section describes how the dataset is engineered.

3.2.2.2 Feature Engineering and Feature Scaling

The first step in making the datasets easier to analyze, or engineer, is to transform any categorical data. The only categorical attribute preserved from the USDA dataset is FoodGroup. To transform this attribute into a numerical format, Scikit-Learn's One Hot Encoding is used. This preprocessing method creates a binary column for each category and returns a sparse matrix, where in the matrix the value is 1 if an instance is that category and otherwise 0 (Pedregosa et al., 2011).

The next step in feature engineering is to aggregate the two datasets. Each instance's subjective data is based on its' components, meaning the survey values are added dependent on the food itself and a series of conditional statements. There are two new attributes added to the dataset, 'weightCauses' and 'whatsHealthy'. For example, the 'weightCauses' value first checks if the food entry has a high amount of sugar. If it does, it is assigned the value of 0.22, the ratio of participants that ranked high sugar levels as the most important in weight gain. If it does not, the next ranked component from the survey is checked, and so on. The 'whatsHealthy' values are assigned with similar conditional statements. The combination process results in a dataset with the same number of entries as the USDA dataset, 8,617, but now with 42 attributes for each instance.

After aggregation, the last step is to determine the output labels for each food instance. Since the USDA dataset does not provide a health label, it is necessary to create this new feature. This is done through the unsupervised clustering algorithm K-Means Clustering (Pedregosa et al., 2011), that clusters data points together based on certain similarities that minimize each cluster's sum of squares. Before applying the algorithm, the full dataset is normalized with Scikit-Learn's Normalizer. It is then fit to the algorithm with a goal of two classes (Healthy and Unhealthy). This results in a division of classes with counts of 3,540 and 5,078. In order to determine the label of each class, the two groups are analyzed by food group. It is found that class 0 is Unhealthy, while class 1 is Healthy. For example, class 0 contains 315 instances of Sweets, while class 1 has 47. Adding this new feature results in a final dataset of 43 attributes.

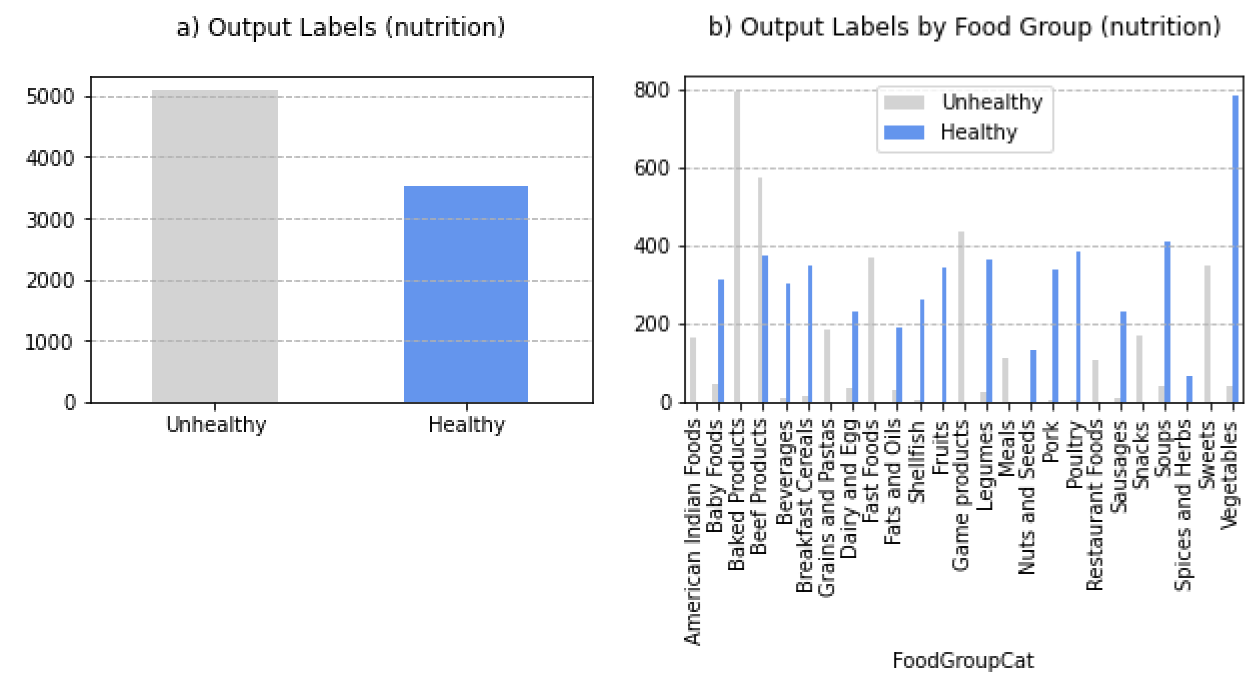

Figure 3.2: Overview of nutrition dataset. a) shows there are more instances of unhealthy in nutrition dataset, b) shows label counts by Food Group in the dataset.

3.2.2.3 Overview of complete nutrition dataset

The final full dataset, now referred to as the nutrition dataset, has 43 attributes and 8,617 instances. Of these instances, 5,078 are of the class Healthy (1) and 3,540 are of the class Unhealthy (0).

The nutrition dataset is now ready for training and testing. It is split into a train and test set, with the dataset shuffled. The train set is used in training the model, which holds 95% of the data, while the test set holds the instances the model will predict on, 5% of the full data. The training set has 8,187 instances and the test set has 431 instances. The training set will be further divided into a train and test validation set during decision making.

CHAPTER 4

MODEL ENVIRONMENT AND IMPLEMENTATION

This chapter details the environment and implementation of the proposed computational model that follows the RAFT cognitive process. First, the model is contextualized within each domain, defining the testing setting and parameters. Then the execution of the model in each domain is outlined, intimating where the domains diverge.

4.1 Environment

The proposed model is required to answer two different questions due to the two problem domains. But the questions are formulated in the same way, i.e. with a binary output.

For the COVID-19 domain, the model is asked to predict if a certain range of dates will have an associated decrease (0) or increase (1) in cases or deaths. This generates two different datasets to analyze and compare, cases and deaths. For both sets, the testing environment is predefined by start date, range of time of information given, range of time to predict over, and range of time between decisions. Table 4.1 shows the possible values for each of these variables. For each of the fifteen start dates, there are four amounts of information given at T1, between one and four weeks. Similarly, for each start date and each amount of time given at T1, there are four different time periods to predict over, between one and four weeks, and the amount of time between decisions is between one and twenty weeks.

In the nutrition domain, the model is asked to determine whether a single food is Unhealthy (0) or Healthy (1). The testing parameters are predefined in the same way as the COVID-19 domain and are shown in table 4.2. The amount of information given at T1 is an interval of the data, the information added at T2 is an interval of the

Table 4.1: Testing environment for COVID-19.

Parameter	Possible values
Start dates (at T1)	3/3/20 4/3/20 5/3/20 6/3/20 7/3/20 8/3/20 9/3/20 10/3/20 11/3/20 12/3/20 1/3/21 2/3/21 3/3/21
Information given at T1 (weeks)	4 6 8 10
Prediction time (same for T1 and T3) (weeks)	1 2 3 4
Time between decisions (weeks)	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Table 4.2: Testing environment for nutrition.

Parameter	Possible values
Given information at T1 (in intervals)	0.1 0.25 0.4 0.5 0.65 0.75 0.9
Information added at T2 (in intervals)	0.1 0.25 0.4 0.5 0.65 0.75 0.9
Time between decisions (weeks)	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

remaining data, and the time between decisions is between one and twenty weeks.

The main difference between the two domains is that information in nutrition is not dependent on time. While the COVID-19 data must be given in order of date, the nutrition data is shuffled. The nutrition testing parameters are in the form of intervals instead of weeks of time.

The following sections walk through the three time steps of the computational model's process for each scenario, highlighting where there are differences in implementing the proposed model.

4.2 Implementation for COVID-19

4.2.1 Time 1

In each COVID-19 testing scenario, the model is given a start date and end date that delineates the given information. From this range of dates, the model is asked to predict if cases or deaths will increase or decrease over each of the prediction ranges, one to four weeks. First a correlation matrix is calculated, then the Take the Best heuristic steps are applied by (1) searching the cues, (2) stopping at the cue with highest discrimination, making a (3) decision rule about cue value, and calculating a (4) confidence value in the chosen cue.

The next step in the RAFT model is to make the prediction, which is achieved through the machine learning classifiers. The T1 given information is split into train-test sets with a test size of 30%, normalized, and then fit to the three classifiers: K Nearest Neighbors (KNN) Classifier, Decision Tree (DT) Classifier, and Random Forest (RF) Classifier. All hyperparameters were determined using Scikit-Learn's Grid Search (Pedregosa et al., 2011) and can be found in Appendix B. After each model is trained and tested, the classifiers and their associated accuracies are stored.

The model then selects the classifier with the highest accuracy. Once the highest ranked classifier is selected, the model then uses that classifier to predict the output labels for each day over the range of prediction dates, where each day corresponds to a value of 1 (increase) or 0 (decrease). The predictions are then averaged. The decision value is computed as the logistic, so that it is between 0 and 1, of the confidence value, classifier accuracy, and prediction value. This decision value is compared to the output threshold value of 0.5 that translates to a binary output label of 0 (decrease) or 1 (increase).

4.2.2 Time 2 and 3

After the model’s decision, a new date and information is given at T2 dependent on the time between decisions, and then at T3 the model is asked to remember the decision made in T1. The original response is deliberated on following the RAFT cognitive process, where either the model is able to access its’ previous decision in memory or will have to reconstruct its' decision by repeating the steps from T1. Here, time between decisions is introduced as a random variable in weeks, with a range of 0 to 20 weeks. Whether the model is able or not to access its' original response is determined by a random choice, i.e. half of the time it can remember and half of the time it must reconstruct the original response. A random choice is used here in order to get balanced occurrences of recall and reconstruction for analysis.

If it is determined that the model is able to remember at T3, a new confidence and decision value are calculated with time between decisions. The new decision value is compared to the output threshold and the resulting binary output label is the T3 decision.

If the model is unable to remember the original response, it is necessary to reconstruct the response. This is done following the same steps from T1, but now the given information used to make the decision includes T1 and T2. The T3 confidence is calculated, the given information is split and normalized, all classifiers are trained and tested, and a decision value is found using the best classifier, confidence value, and time between decisions. From this decision value, a choice is made by using the threshold value.

4.3 Implementation for Nutrition

Moving on from COVID-19, now the implementation for nutrition is described. Both domains are described because there are a few differences in implementation due to the separate contexts. The implementation of the model for nutrition is not described in the same detail as for COVID-19 since much of it is repetitive. Instead, the differences are mentioned.

4.3.1 Time 1

At T1, the model is given a rate value that corresponds to the amount of information given from the training dataset. In the preprocessing step, the nutrition data was split into a train-test set with a test size of 5%. The given information is taken from the training dataset using the Pandas sample() method, proportional to the rate given. The sample() method returns a random sample of the size given. A single test instance is sampled from the test dataset in the same way. The TTB process is applied to the information given at T1, following the same implementation for COVID-19. The given information is split into validation train and test sets with a test size of 30%, then applied to the same three classifiers. The hyperparameters are detailed in Appendix B. The classifiers and their accuracies are stored. Using the best classifier, a prediction is made on the single test instance, given in the form of probabilities for each class. The prediction value, the confidence value, and classifier accuracy are then calculated the same way as the COVID-19 decision value. The decision value is applied to the threshold value that indicates the class that is chosen (healthy or unhealthy).

4.3.2 Time 2 and 3

After the original response is made, a random amount of time is introduced. New information is added, where the amount of information added is given from the T2 rates. The added information is a portion of the remaining training set of data after divided at T1, and again is found through the sample method. This is one of the main implementation differences, since the nutrition dataset does not add new information that is proportional to the time between decisions.

If it is determined that the model is able to remember after the given time, a new decision value and decision are calculated in the same way as described above.

If the model is unable to remember the values from T1, it reconstructs the original response using the steps from T1, in the same process as the COVID-19 implementation.

CHAPTER 5

PREDICTIONS

This chapter introduces the predictions before testing is performed. There are three groups of predictions, those only concerning the COVID-19 domain, those only concerning the nutrition domain, and those connecting the two.

5.1 COVID-19

Within the COVID-19 scenarios, there are four kinds of predictions. The first three outline different measurements of hindsight bias: through the change in hindsight bias (HBC) equation introduced in chapter 2, by confidence values of T1 and T3 decisions, and by comparison of T1 and T3 decisions to the correct response (CR).

Across the COVID-19 domain, it is predicted that when decision-making is unimpaired, hindsight bias will be low. The three assumed sources of impairment for COVID-19 are the time given at T1, the prediction size, and the time between decisions. The first three categories of predictions detailed below show the results in the context of each source of impairment. The fourth category compares the death and case testing scenarios in the COVID-19 domain.

The first category of predictions concerns the three sources of assumed impairment as measured by the HBC equation. The predictions are defined and explained below, for both cases and deaths.

> COVID-19 (1.a): *As the amount of time given at T1 increases, HBC averages will decrease.* Having less information to make an informed decision is an impairment, where increasing the amount of information will lead to less HBC.
>
> COVID-19 (1.b): *As the prediction size increases, HBC averages will increase*. Here, predicting over a larger amount of time is expected to reflect that it is harder to accurately decide, leading to more HBC.
>
> COVID-19 (1.c): *As the time between decisions increases, HBC averages will increase.* With more time between decisions, the ability to accurately remember or reconstruct the original response will become more impaired.

The second category of predictions for COVID-19 uses confidence values of T1 and T3 decisions as a measure of hindsight bias. This category's predictions are defined and explained below, for both cases and deaths. Overall, it is expected that *T3 will be associated with lower levels of confidence than T1.* Since it is expected that T3 decisions are made in the face of more impairment, T3 decisions will have lower confidence values.

> COVID-19 (2.a): *As the given time at T1 increases, confidence values will increase*. Decisions are more impaired with less information given, so the confidence in decisions will increase with more time given.

> COVID-19 (2.b): *As the prediction size increases, confidence values will decrease.* With more time to predict over, it is expected that decisions will become more impaired.

> COVID-19 (2.c): *As the time between decisions increases, confidence values will decrease.* With more time between decisions, it is expected there will be more trouble accurately recalling or reconstructing the original response, leading to lower confidence values.

The third category of predictions evaluates how the COVID-19 model performed in deciding the correct response (CR). Overall, it is predicted that the *T3 decisions will be closer to the CR than T1 decisions*. The relevant predictions are defined and explained below, for both cases and deaths.

> COVID-19 (3.a): *As the time given at T1 increases, the amount of CR will increase.* Less given time is an impairment that will be associated with less CR.

> COVID-19 (3.b): *As the prediction size increases, the amount of CR will decrease.* By predicting over larger amounts of time, the ability to decide the CR will decrease.

> COVID-19 (3.c): *As the time between decisions increases, the amount CR will increase.* It is predicted that after more time, there will be higher impairment in accurately recollecting or reconstructing the original response. With more time, decisions will move towards the CR.

The fourth COVID-19 category compares the case and death testing scenarios within the COVID-19 domain. It is predicted that the COVID-19 death results will be associated with higher hindsight bias. Since the number of cases throughout the dataset is far greater and more variable than the number of deaths, and since deaths are more subjectively consequential, as conveyed in the survey data, it is expected the results will reveal greater hindsight bias for deaths than for cases.

COVID-19 (4): *Greater hindsight bias will be observed with respect to deaths than cases.*

5.2 Nutrition

The nutrition setting's predictions follow the same first three categories as in the COVID-19 domain, but does not include the fourth category. Within the three nutrition categories, there are two differences from their COVID-19 parallels.

First, the nutrition setting does not use the concept of time in the same way. The given information at T1 is in the form of rates instead of weeks. Subsequently, the amount of information at T2 is not determined by weeks after the original decision, but is instead in the form rates. The second difference is that the nutrition setting does not have a variable for prediction size, instead making decisions for one instance. As a results of these two differences, the assumed impairment of prediction size in the COVID-19 domain is replaced by the T2 information rate in the nutrition domain. It is predicted that as the T2 information rate increases, impairment will increase.

For the sake of comprehension and readability, even though time is not relevant in the nutrition dataset, the terminology of time between decisions is retained in the nutrition domain. Here, time between decisions should be understood as the amount of added uncertainty between decisions.

The first category of nutrition predictions measures hindsight bias through the HBC equation. These predictions are defined and explained below.

Nutrition (1.a): *As the given rate at T1 increases, HBC averages will decrease.* It is expected that less information is an impairment, so HBC averages will decrease with more information.

Nutrition (1.b): *As the given rate at T2 increases, HBC averages will decrease*. It is expected that less information is an impairment, so HBC averages will decrease with more T2 information given.

Nutrition (1.c): *As the time between decisions increases, HBC averages will increase.* More time between decisions will impair accurately remembering or reconstructing the original decision.

The second category of nutrition predictions measures hindsight bias through confidence values. Overall, it is expected that *T3 decisions will have lower levels of confidence than T1.* This category's predictions are defined and explained below.

Nutrition (2.a): *As the given rate at T1 increases, confidence values will increase.* Decisions are impaired with less information given, so the confidence in decisions will increase as the information rate increases.

Nutrition (2.b): *As the given rate at T2 increases, confidence values will increase at T3*. With more information at T3 deliberation, confidence values will become less impaired.

Nutrition (2.c): *As time between decisions increases, confidence values will decrease.* With increasing impairment, confidence will decrease.

The third category of nutrition predictions measures hindsight bias by examining the amount of correct responses (CR) at T1 and T3. These predictions are defined and explained below.

Nutrition (3.a): *As the given rate at T1 increases, the amount of CR will increase.* With less impairment, the amount of CR will increase.

Nutrition (3.b): *As the given rate at T2 increases, the amount of CR will increase.* With less impairment at T3 deliberation, there will be more CR at T3.

Nutrition (3.c): *As the time between decisions increases, the amount of CR will increase.* With more time between, T3 decisions will move towards the CR.

5.3 Combined COVID-19 and Nutrition

The third grouping of predictions compares the two testing domains. It is predicted that the COVID-19 results will display higher levels of hindsight bias. This prediction is made for three reasons. First, that COVID-19 is a more novel and socially charged domain than nutrition, meaning the difference between the original response and correct response is weighted more in the COVID-19 domain. Second, the COVID-19 scenario predicts over a range of instances (dates), while the nutrition scenario must predict over one instance (a food). Having one prediction instance makes the problem simpler for the nutrition scenario. Third, the nutrition dataset has more training instances than the COVID-19 dataset. It is predicted the nutrition implementation's recalled or reconstructed decisions will be more influenced by the degree of information available than COVID-19, resulting in more hindsight bias.

> Combined (1): *Greater hindsight bias will be observed with respect to COVID-19 than nutrition.*

CHAPTER 6

RESULTS

The results chapter first contextualizes COVID-19 with its predictions, followed by nutrition, and then both domains combined.

6.1 COVID-19

This section describes the results in the context of the aforementioned COVID-19 predictions. In COVID-19 testing, there were a total of thirteen start dates, four options of amount of T1 given information, a range of one to four weeks for the prediction date, and twenty possible times between decisions. Each start date was tested five times, resulting in a total of 4,160 test instances for both cases and deaths. Other general results and statistical descriptions are given in Appendix C.

COVID-19 results are given for both cases and deaths within each prediction category. There are assumed to be three sources of impairment that impact hindsight bias: time given at T1, prediction size, and time between decisions. The first prediction category measures hindsight bias with the HBC equation, the second uses confidence values, and the third compares T1 and T3 decisions to the correct responses (CR). The fourth category of COVID-19 predictions compares the results of cases and deaths.

6.1.1 COVID-19 (1) measuring hindsight bias with HBC equation

The first prediction category utilizes the HBC equation to measure hindsight bias.

1.a As the time given at T1 increases, HBC averages will decrease

Time given at T1 is in the form of four, six, eight, or ten weeks, and it is predicted that as time increases, the average HBC will decrease. Cases results are associated with an overall decrease in HBC from four to ten weeks, but show eight weeks to have the highest HBC average (0.129) in figure 6.1. The results for deaths show that the average HBC decreases from four to ten weeks (figure 6.1). For both cases and deaths, time given at T1 appears as expected to be a source of hindsight bias as measured by HBC.

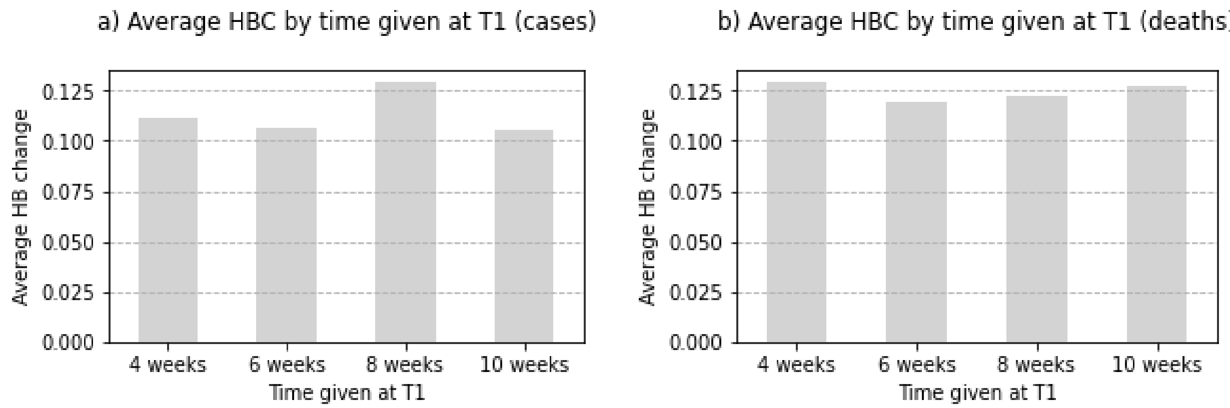

Figure 6.1: COVID-19 (1.a) prediction results of average HBC by time given at T1. a) shows the results for cases, with an overall decrease from 4 weeks to 10, b) shows the results for deaths, with HBC averages decreasing overall from 4 weeks to 10.

1.b As prediction size increases, HBC averages will increase

It is expected that HBC averages will increase as prediction size grows since deliberation will have higher impairment. Figure 6.2 shows that for cases, the average HBC increases with prediction size. Death results, while having an overall increase in HBC from one to four weeks, do not show a significant difference in HBC averages over prediction size (figure 6.2). HBC averages increase with prediction size for cases and deaths, showing prediction size to function as expected as a source of impairment when measured by HBC.

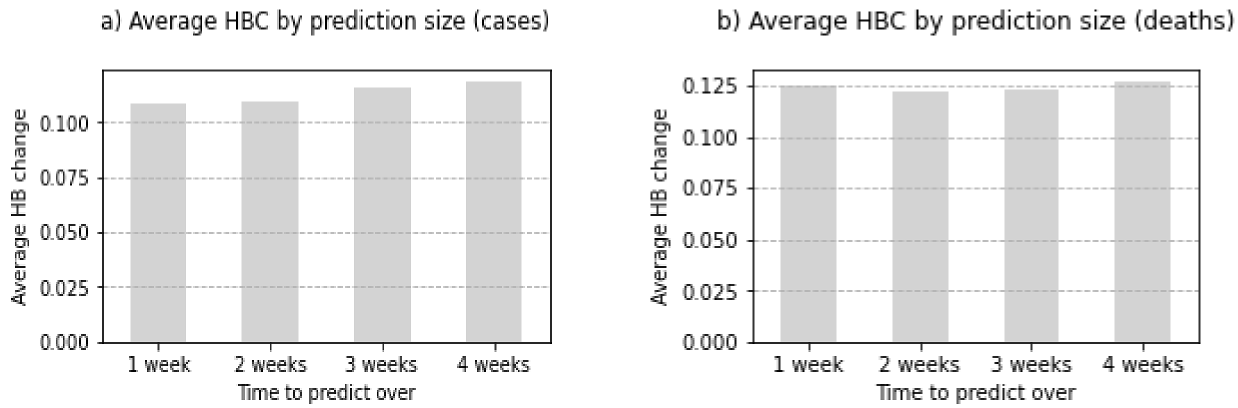

Figure 6.2: COVID-19 (1.b) prediction results of average HBC by prediction time. a) shows the results for cases, with an increase in HBC with prediction size, b) shows an overall increase in HBC from 1 to 4 weeks.

1.c As time between decisions increases, HBC averages will increase

With time between decisions increasing, it is predicted that there will be greater HBC averages. The cases testing performs as expected, where the HBC averages increase (figure 6.3). Similarly, the average HBC increases with time between decisions for deaths, with each value being larger than those for cases (figure 6.3). Results show time between decisions to be a strong source of impairment for cases and deaths as measured by HBC.

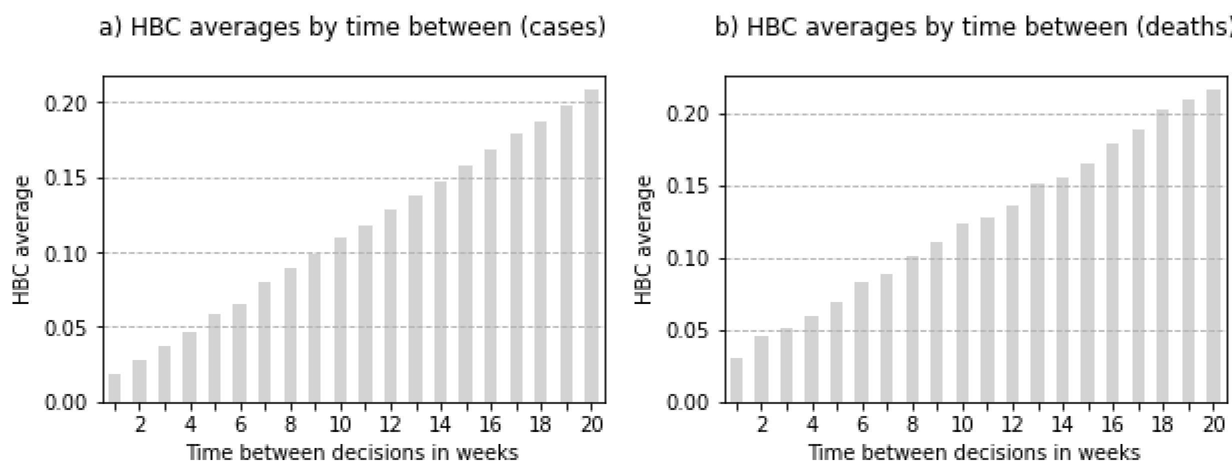

Figure 6.3: COVID-19 (1.c) prediction results of average HBC by time between decisions. a) shows the results for cases, with almost proportional increase of HBC average with time between decisions, b) shows the same trend for deaths as cases, with higher values at each time between decisions.

In summary for the COVID-19 (1) predictions, the given time at T1 and prediction size appear as expected as sources of impairment, but time between decisions is revealed to be a strong source of impairment for hindsight bias for cases and deaths in the COVID-19 domain.

6.1.2 COVID-19 (2) measuring hindsight bias with confidence values

This prediction category examines the confidence values of T1 and T3 decisions as a measurement of hindsight bias. Overall, it is predicted that *T3 will have lower levels of confidence than T1* for the COVID-19 domain, reflecting that as impairment increases, confidence values will decrease. Table 6.1 shows, for both cases and deaths, confidence values decrease from T1 to T3, with deaths changing more significantly (cases: 4.6% difference, deaths: 29.1% difference).

Table 6.1: Confidence values at T1 and T3 for COVID-19 cases and deaths. The percent difference for deaths is greater than that for cases.

	Cases			Deaths		
	T1 confidence	T3 confidence	% difference	T1 confidence	T3 confidence	% difference
mean	0.7929	0.7561	-4.6499	0.7968	0.5653	-29.0534
std	0.0596	0.0507	--	0.0523	0.1329	--

2.a As the rate of information at T1 increases, confidence values will increase

It is predicted that less information at T1 is an impairment, where increasing time given at T1 will lead to an increase in confidence values. Table 6.2 shows that confidence values decrease with time for cases T1 and T3, and for deaths T3. T1 confidence values for deaths increase with T1 time given. For both cases and deaths, the confidence values at T3 are less than those at T1. From these results, T1 given information performs contradictory to the predictions, with confidence values overall decreasing for cases and deaths by time given at T1.

Table 6.2: COVID-19 (2.a) prediction results of T1 and T3 confidence values for cases and deaths by time given at T1.

		Cases					Deaths				
		T1 confidence		T3 confidence		% difference	T1 confidence		T3 confidence		% difference
		mean	**std**	**mean**	**std**		**mean**	**std**	**mean**	**std**	
T1 given time	**4**	0.7987	0.0808	0.7661	0.049505	-4.08	0.6378	0.0514	0.5729	0.1310	-10.19
	6	0.7968	0.0520	0.7600	0.050501	-4.63	0.6479	0.0541	0.5586	0.1304	-13.79
	8	0.7924	0.0541	0.7556	0.050198	-4.65	0.6697	0.0572	0.5543	0.1309	-17.23
	10	0.7838	0.0440	0.7426	0.049925	-5.26	0.6987	0.0788	0.5531	0.1309	-20.83

2.b As prediction size increases, confidence values will decrease

It is expected that as prediction size increases, confidence values at T1 and T3 will decrease due to greater impairment. The results for cases show an increase in confidence values with prediction size at T1 and T3, while results for deaths show the expected decrease in confidence values at T1 and T3 (table 6.3). Prediction size functions as expected only in the context of COVID-19 deaths. Prediction size is revealed to be contradictory to the predictions, with confidence values increasing with prediction size for cases and deaths.

Table 6.3: COVID-19 (2.b) prediction results of T1 and T3 confidence values for cases and deaths by time to predict over. The percent differences for deaths are greater than cases.

		Cases					Deaths				
		T1 confidence		T3 confidence		% difference	T1 confidence		T3 confidence		% difference
		mean	**std**	**mean**	**std**		**mean**	**std**	**mean**	**std**	
Time to predict over	**1**	0.7899	0.0496	0.7559	0.0502	-4.31	0.6793	0.0596	0.5637	0.1304	-17.02
	2	0.7902	0.0592	0.7547	0.0504	-4.49	0.6693	0.0592	0.5561	0.1295	-16.91
	3	0.7909	0.0660	0.7562	0.0526	-4.39	0.6699	0.0606	0.5600	0.1303	-16.41
	4	0.7929	0.0696	0.7574	0.0499	-4.48	0.6479	0.0575	0.5596	0.1321	-13.64

2.c As time between decisions increase, confidence values will decrease

In examining time between decisions and confidence values, it is predicted that confidence values decrease with an increase in time at T3, since T1 decisions are made independently of time between decisions. This trend occurs in the results for both cases and deaths. The change in confidence values from one week to twenty is more significant for deaths (53.5% difference) than for cases (8.5% difference). The results show that time between decisions is a strong source of impairment for both cases and deaths as measure by confidence values.

Table 6.4: COVID-19 (2.c) prediction results of T1 and T3 confidence values by time between decisions for cases and deaths. The percent differences are greater for deaths than cases.

	Cases					Deaths				
	T1 confidence		T3 confidence		% difference	T1 confidence		T3 confidence		% difference
Time between decisions	**mean**	**std**	**mean**	**std**		**mean**	**std**	**mean**	**std**	
1	0.7929	0.0597	0.7855	0.0513	-0.93	0.7729	0.0687	0.7647	0.0555	-1.06
2	0.7929	0.0597	0.7885	0.0502	-0.56	0.7729	0.0687	0.7508	0.0457	-2.86
3	0.7929	0.0597	0.7809	0.0486	-1.52	0.7729	0.0687	0.7242	0.0449	-6.31
4	0.7929	0.0597	0.7776	0.0495	-1.93	0.7729	0.0687	0.7016	0.0420	-9.23
5	0.7929	0.0597	0.7753	0.0473	-2.23	0.7729	0.0687	0.6800	0.0410	-12.03
6	0.7929	0.0597	0.7761	0.0448	-2.13	0.7729	0.0687	0.6585	0.0375	-14.81
7	0.7929	0.0597	0.7713	0.0429	-2.73	0.7729	0.0687	0.6368	0.0391	-17.61
8	0.7929	0.0597	0.7711	0.0439	-2.75	0.7729	0.0687	0.6061	0.0370	-21.59
9	0.7929	0.0597	0.7675	0.0432	-3.21	0.7729	0.0687	0.5918	0.0338	-23.44
10	0.7929	0.0597	0.7596	0.0415	-4.20	0.7729	0.0687	0.5727	0.0309	-25.91
11	0.7929	0.0597	0.7534	0.0428	-4.99	0.7729	0.0687	0.5478	0.0302	-29.13
12	0.7929	0.0597	0.7514	0.0375	-5.23	0.7729	0.0687	0.5248	0.0289	-32.10
13	0.7929	0.0597	0.7450	0.0422	-6.05	0.7729	0.0687	0.5098	0.0260	-34.05
14	0.7929	0.0597	0.7456	0.0478	-5.97	0.7729	0.0687	0.4802	0.0317	-37.87
15	0.7929	0.0597	0.7362	0.0495	-7.16	0.7729	0.0687	0.4579	0.0279	-40.76
16	0.7929	0.0597	0.7354	0.0447	-7.25	0.7729	0.0687	0.4384	0.0287	-43.28
17	0.7929	0.0597	0.7321	0.0495	-7.67	0.7729	0.0687	0.4171	0.0278	-46.03
18	0.7929	0.0597	0.7270	0.0463	-8.31	0.7729	0.0687	0.3944	0.0262	-48.98
19	0.7929	0.0597	0.7204	0.0504	-9.15	0.7729	0.0687	0.3803	0.0257	-50.80
20	0.7929	0.0597	0.7187	0.0462	-9.37	0.7729	0.0687	0.3555	0.0253	-54.00

Examining hindsight bias through confidence levels reveals both time given at T1 and time to predict over to be contradictory sources of impairment for hindsight bias when measured by confidence values. Time between decisions is found to be a strong source of impairment for both cases and deaths. Overall, the percent differences in confidence values from T1 to T3 for deaths are larger than for cases.

6.1.3 COVID-19 (3) measuring hindsight bias with correct responses

This COVID-19 prediction category analyzes T1 and T3 decisions compared to the correct responses (CR). Overall, it is predicted that the *T3 decisions will be closer to the CR than T1 decisions*. Figure 6.4 shows this trend for cases, where the amount of correct responses is larger at T3 than T1, moving from 88.94% correct to 95.12% (6.95% difference). Death results show the opposite, where T3 has less correct responses than T1, moving from 88.46% correct decisions to 76.25% (-13.98% difference). While the previous measures have shown hindsight bias occurs for deaths, comparing the decisions to the CR reveals that T3 decisions move towards incorrect decisions. This means that death results have less accurate hindsight bias than cases.

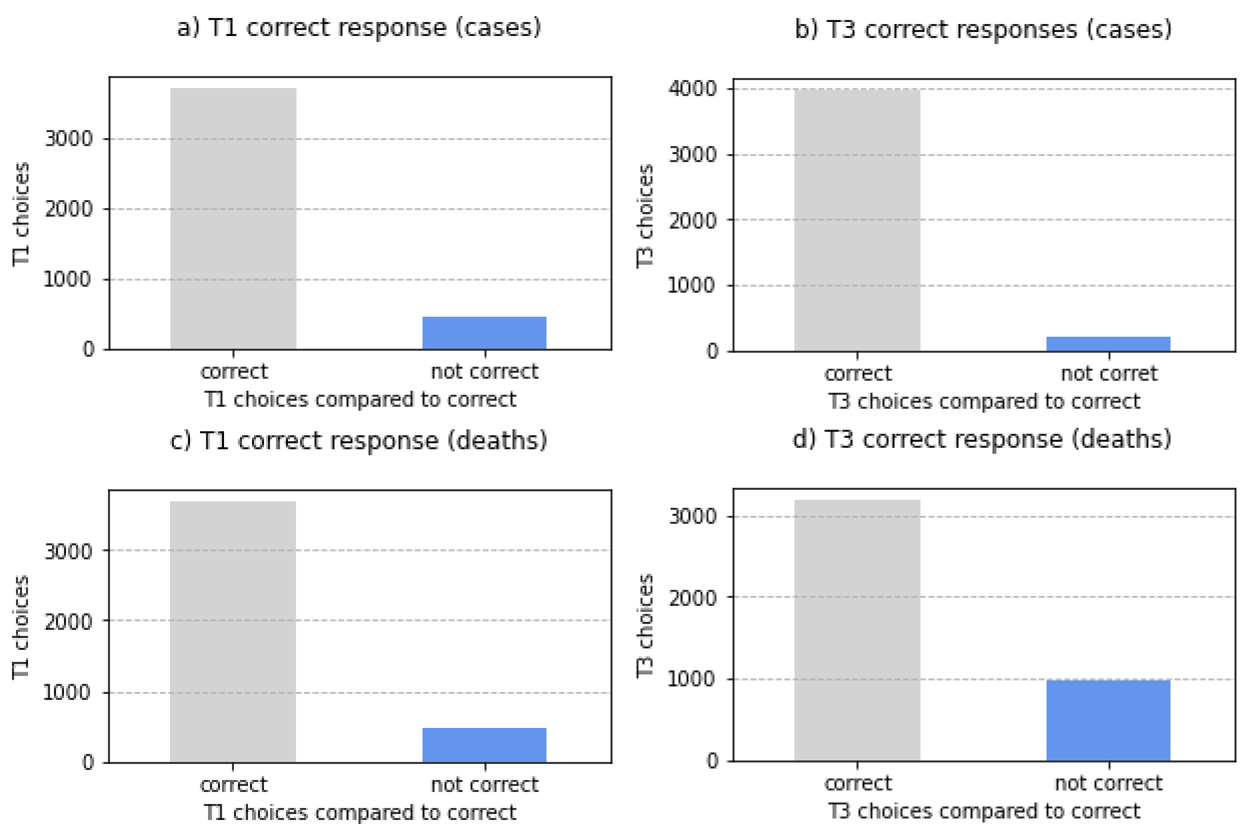

Figure 6.4: CR at T1 and T3 for COVID-19 cases and deaths. a) shows the choices made at T1 for cases, b) shows the number of correct decisions increases at T3 for cases, c) shows the choices made at T1 for deaths, d) shows the number of correct decisions decreases at T3 for deaths.

3.a As given time increases, the amount of CR will increase

With increased given information at T1, it is predicted the amount of CR will increase. Cases show this trend for both T1 and T3, and show that T3 has higher amounts of CR (figure 6.5). For deaths, CR decreases with time given, but T1 has more CR than T3. Time given at T1 is a strong source of impairment for cases as measured by CR, but acts contradictory for deaths.

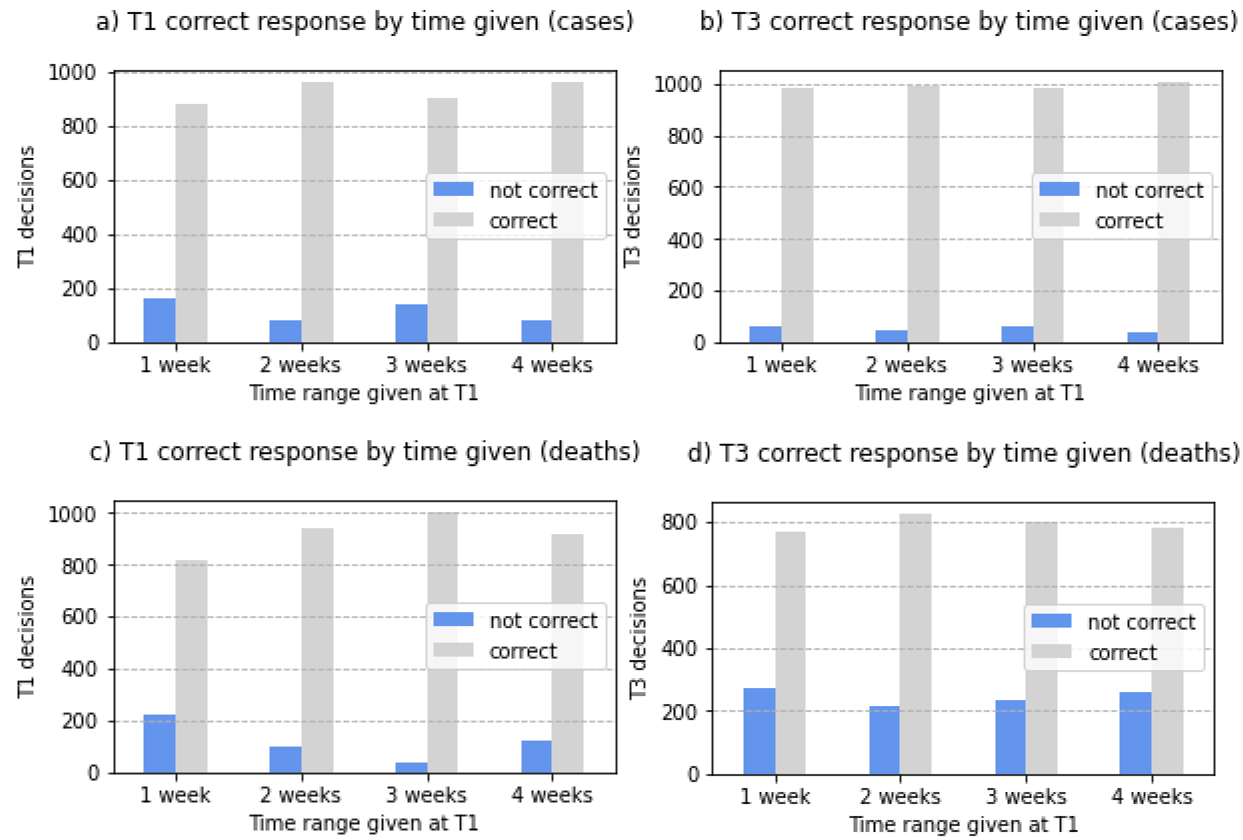

Figure 6.5: COVID-19 (3.a) prediction results of CR at T1 and T3 by time given at T1. a) shows the CR for cases at T1, b) shows CR for cases at T3, with more correct than at T1, c) shows the CR for deaths at T1, d) shows CR for deaths at T3, with less correct than at T1.

3.b As prediction size increases, CR will decrease

It is expected that as prediction size grows, the amount of CR will decrease. CR are lowest at prediction sizes of one and four weeks for cases but does not have a clear trend. Figure 6.6 shows that for deaths, as prediction size increases, the number of CR decreases as expected. Overall, prediction size is a strong indicator of hindsight bias for deaths but is inconsequential for cases.

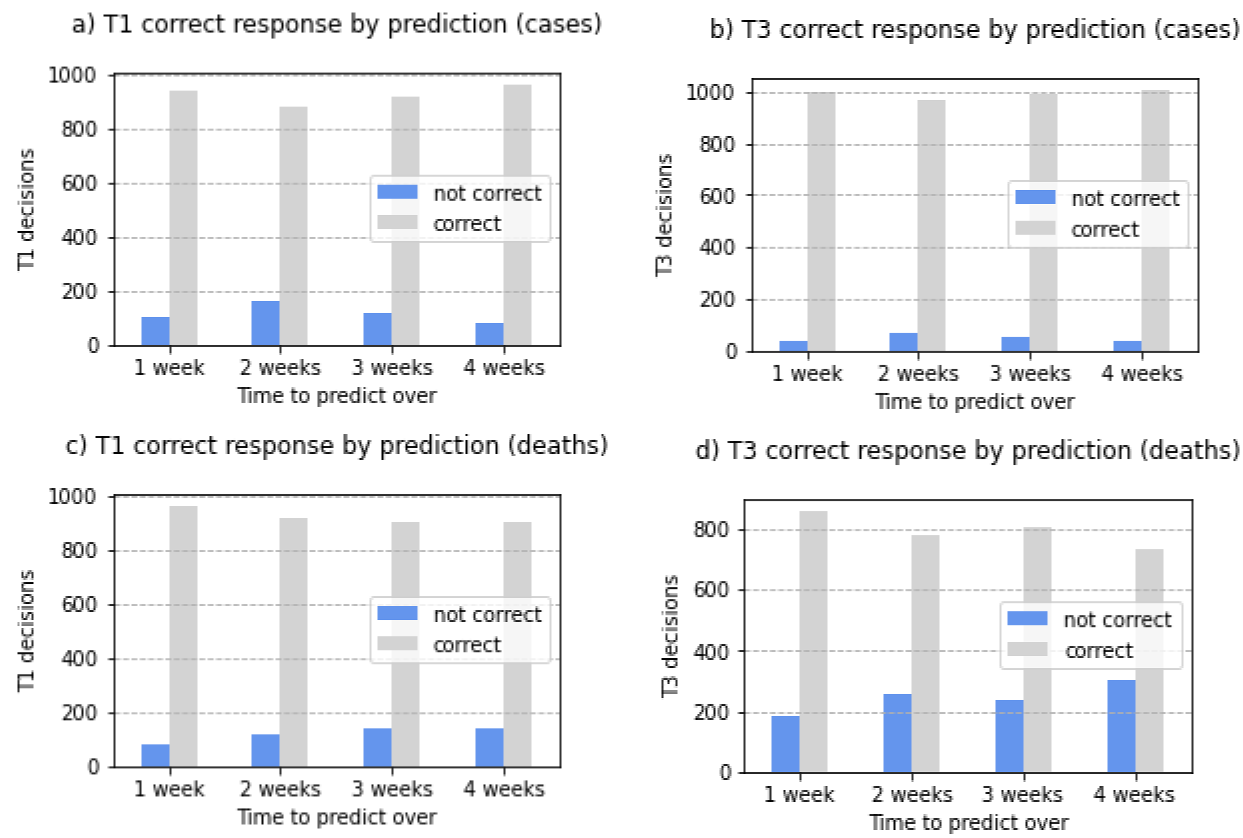

Figure 6.6: COVID-19 (3.b) prediction results of CR at T1 and T3 by prediction size. a) shows the CR for cases at T1, b) shows the CR for cases at T3, with more correct than at T1, c) shows the CR for deaths at T1, d) shows CR for deaths at T3, with less correct than at T1.

3.c As time between decisions increases, CR will increase

It is predicted that as time between decisions increases, the more CR are made at T3, since T1 decisions are independent of time between decisions. For cases, the number of incorrect decisions steadily drops until 17 weeks between decisions, where it drops significantly (figure 6.7). Deaths show a similar trend, with the number of incorrect decisions dropping steadily until 18 weeks between decisions. Time between decisions as an impairment functions as expected for both cases and deaths.

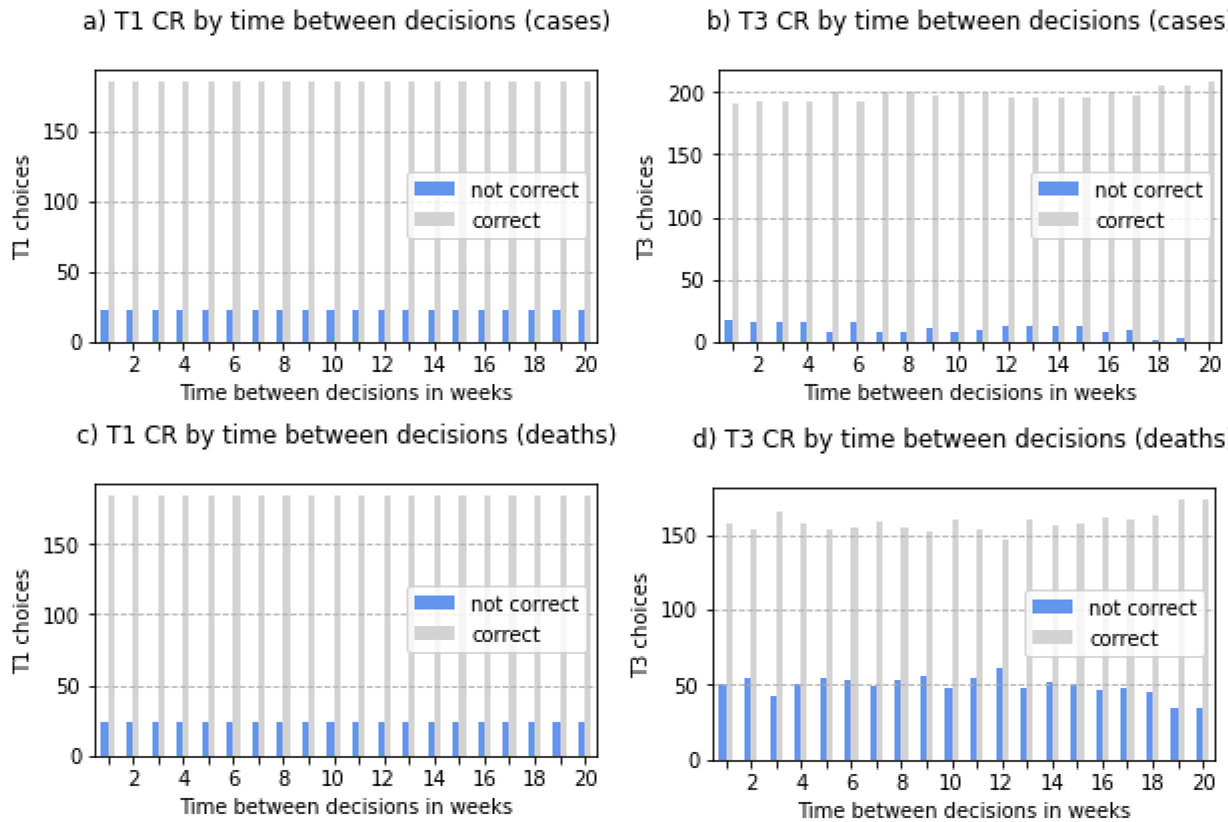

Figure 6.7: COVID-19 (3.c) prediction results of CR at T1 and T3 by time between decisions. a) shows CR for cases at T1, b) shows CR for cases at T3, where CR increase with time, c) shows CR for deaths at T1, d) shows CR for deaths at T3, where CR increase with time.

Within category three that examines hindsight bias through CR, it is shown that overall, cases have a more accurate shift towards the CR than deaths (figure 6.4), exemplified by there being more amounts of CR for cases in the context of each source of impairment. Prediction size is shown to be a weak indicator of hindsight bias for cases, while time given at T1 and time between decisions are strong sources for both cases and deaths.

6.1.4 COVID-19 (4): Death results will have higher hindsight bias than cases

Overall, it is predicted that death results will display more hindsight bias than cases. This expectation is confirmed in each prediction category. In prediction category one, death results reveal higher HBC averages for each source of impairment (figures 6.1-6.3). For category two, the percent of difference between T1 and T3 confidence levels is greater for deaths than cases (cases: 4.65%, deaths: 29.05%, table 6.1). In the third prediction category, the death results reveal higher percent differences in T1 to T3 CR (cases: 6.95%, deaths: 13.98), even though this difference leads to deaths having more incorrect responses than cases. Each category of COVID-19 predictions show deaths to be associated with greater hindsight bias than cases.

6.2 Nutrition

This section explores the results of nutrition testing in the context of the relevant predictions. For nutrition, there are a total of nine rates of information given at T1, nine rates of information added at T2, and twenty possible values for time between decisions. Since nutrition does not have an input for prediction size, the whole process is tested five times to get comparable amounts of test instances. This results in a total of 4,900 test instances for nutrition. Other general results and statistics are given in Appendix C.

The results section for nutrition follows the same format as for COVID-19, without the fourth prediction category and with the T2 given rate replacing prediction size. The three sources of assumed impairment are: rate of information given at T1, rate of information given at T2, and time between decisions. The first category of predictions measures hindsight bias with the HBC equation, the second category uses confidence values, and the third compares decisions to the correct responses (CR).

6.2.1 Nutrition (1) measuring hindsight bias with HBC equation

1.a As rate given at T1 increases, HBC averages will decrease

In nutrition, it is expected that as the given rate at T1 increases, the average HBC will decrease. The results show the average HBC increasing instead as the given rate of information increases (figure 6.8). The T1 given rate is found to be a contradictory source of impairment for hindsight bias as measured by HBC.

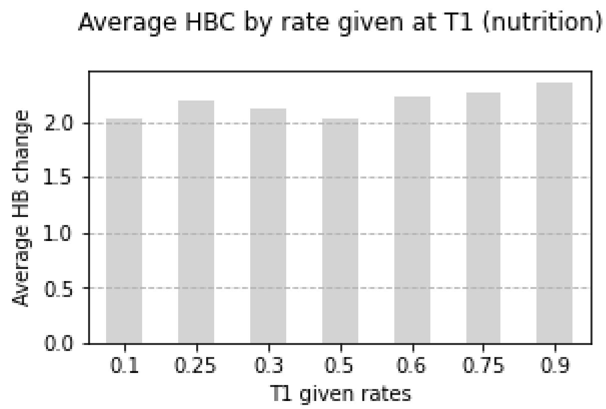

Figure 6.8: Nutrition (1.a) prediction results of average HBC by given rate at T1. Average HBC increases as amount of information given at T1 increases.

1.b As T2 given rate increases, HBC averages will decrease

With the rate of information added at T2, it is predicted that the average HBC will decrease. The HBC averages do not show an overall yet unclear trend of HBC averages increasing with T2 given information (figure 6.9). T2 given information is a contradictory source of impairment for hindsight bias when measures by HBC.

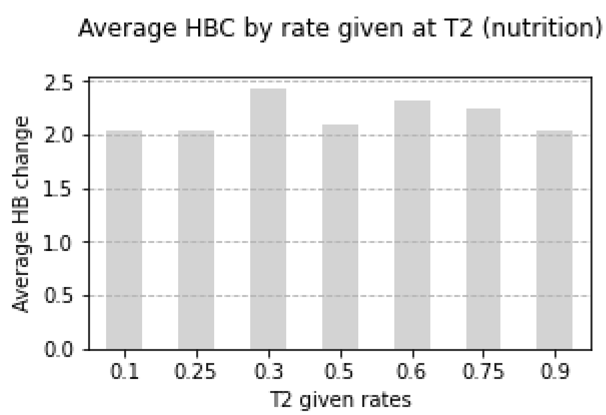

Figure 6.9: Nutrition (1.b) prediction results of average HBC by given rate at T2. Overall HBC averages increase with T2 rate.

1.c As time between decisions increases, HBC averages will increase

With time between decisions increasing, it is predicted that HBC averages will increase. The results show there is an increase in average HBC with time between decisions, as expected (figure 6.10). Time between decisions is a strong source of impairment of hindsight bias as measured by HBC.

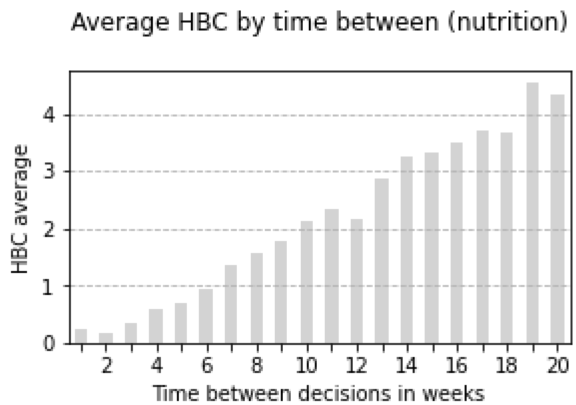

Figure 6.10: Nutrition (1.c) prediction results of average HBC by time between decisions. Average HBC increases with time between decisions.

Within the first category of nutrition predictions, both T1 given rate and T2 given rate are found to be contradictory sources of impairment for hindsight bias for nutrition. Yet time between decisions is a strong source of impairment for nutrition.

6.2.2 Nutrition (2) measuring hindsight bias with confidence values

In the context of the second method for measuring hindsight bias, it is predicted that as impairment increases, confidence values in the nutrition domain will decrease. This means that *T3 decisions will have lower confidence values than decisions made at T1* due to more impairment. Comparison of T1 and T3 confidence values of nutrition do show that T3 confidence is on average lower than T1's, but also reveals there is not much difference in the two (0.0002) (table 6.5). The percent difference in values from T1 to T3 is -0.028.

Table 6.5: Confidence values of T1 and T3 decisions for nutrition. There is slight percent difference from T1 to T3.

	Nutrition		
	T1 confidence	T3 confidence	% difference
mean	0.7737	0.7735	-0.0280
std	0.0077	0.0038	--

2.a As given rate at T1 increases, confidence values will increase

It is predicted that less information at T1 is an impairment, where increasing the rate given at T1 will lead to an increase in confidence values. Table 6.6 shows that confidence values decrease with T1 rate for T1 and there is not clear trend for T3. The percent differences from T1 to T3 decrease for all rates. Given information at T1 is revealed to be a weak source of impairment for hindsight bias as measured by confidence values in the nutrition domain.

Table 6.6: Nutrition (2.a) prediction results of confidence values by T1 given information. The percent differences decrease for all rates.

		Nutrition				
		T1 confidence		T3 confidence		% difference
		mean	**std**	**mean**	**std**	
T1 given rate	**0.1**	0.7740	0.0152	0.7734	0.0062	-0.070
	0.25	0.7744	0.0084	0.7739	0.0047	-0.072
	0.3	0.7739	0.0076	0.7738	0.0031	-0.021
	0.5	0.7734	0.0054	0.7734	0.0499	-0.003
	0.6	0.7735	0.0041	0.7734	0.0027	-0.011
	0.75	0.7734	0.0028	0.7734	0.0020	-0.002
	0.9	0.7735	0.0017	0.7734	0.0012	-0.018

2.b As given rate at T2 increases, confidence values will increase

It is predicted that less information at T1 is an impairment that will lead to an increase in confidence values. Table 6.7 shows there not to be a clear trend for values at T1, but an overall decrease at T3. The percent difference between T1 and T3 show an average decrease from T1 to T3. The percent differences form T1 to T3 decrease on average. The given rate at T1 is revealed to be a weak source of impairment for hindsight bias as measured by confidence values in the nutrition domain.

Table 6.7: Nutrition (2.b) prediction results of confidence values by T2 given information. The percent differences on average decrease for T2 rates.

		Nutrition				
		T1 confidence		T3 confidence		% difference
		mean	**std**	**mean**	**std**	
T2 given rate	**0.1**	0.7736	0.0079	0.7737	0.0061	0.01
	0.25	0.7740	0.0077	0.7737	0.0047	-0.04
	0.3	0.7734	0.0071	0.7736	0.0043	0.02
	0.5	0.7739	0.0077	0.7735	0.0034	-0.04
	0.6	0.7742	0.0071	0.7733	0.0028	-0.11
	0.75	0.7734	0.0078	0.7734	0.0020	0.00
	0.9	0.7737	0.0086	0.7735	0.0012	-0.03

2.c Confidence values will decrease as time between decisions increases

In examining time between decisions and confidence values, it is predicted that confidence values decrease with time between decisions. Results show that confidence values for T1 and T3 overall increase from 1 week to twenty, and the percent differences decrease on average for time between decisions (table 6.8). Time between decisions is found to be a source of impairment as expected, but with small percent changes of confidence values.

Table 6.8: Nutrition (2.c) prediction results of confidence values by time between decisions. Percent differences decrease on average for each time.

		Nutrition				
		T1 confidence		T3 confidence		%
		mean	std	mean	std	difference
Time between decisions	1	0.7735	0.0067	0.7737	0.0036	0.034
	2	0.7738	0.0086	0.7734	0.0039	-0.062
	3	0.7737	0.0074	0.7733	0.0038	-0.060
	4	0.7742	0.0076	0.7737	0.0035	-0.063
	5	0.7736	0.0065	0.7734	0.0038	-0.031
	6	0.7737	0.0080	0.7735	0.0040	-0.021
	7	0.7744	0.0079	0.7737	0.0041	-0.085
	8	0.7743	0.0070	0.7737	0.0038	-0.074
	9	0.7735	0.0083	0.7734	0.0038	-0.022
	10	0.7736	0.0074	0.7736	0.0040	-0.008
	11	0.7737	0.0072	0.7737	0.0041	0.007
	12	0.7736	0.0076	0.7734	0.0037	-0.027
	13	0.7738	0.0075	0.7731	0.0038	-0.090
	14	0.7733	0.0078	0.7737	0.0034	0.043
	15	0.7741	0.0076	0.7736	0.0033	-0.063
	16	0.7733	0.0084	0.7738	0.0041	0.058
	17	0.7727	0.0080	0.7734	0.0038	0.088
	18	0.7744	0.0087	0.7732	0.0045	-0.151
	19	0.7738	0.0079	0.7733	0.0035	-0.058
	20	0.7738	0.0079	0.7740	0.0037	0.021

Examining hindsight bias through confidence levels shows T1 given rates and T2 given rates to be weak sources of impairment for hindsight bias. Time between decisions is found to be a strong indicator of hindsight bias. Overall, nutrition confidence values decrease from T1 to T3, but at low numbers.

6.2.3 Nutrition (3) measuring hindsight bias with correct responses

The third category of predictions analyzes the model results compared to the correct responses (CR). Overall it is predicted that T3 will have more CR than T1. Nutrition results in figure 6.11 show a shift upwards in correctly deciding the CR from an accuracy at T1 of 51.71% to an accuracy of 92.08% at T3 (78.07% difference).

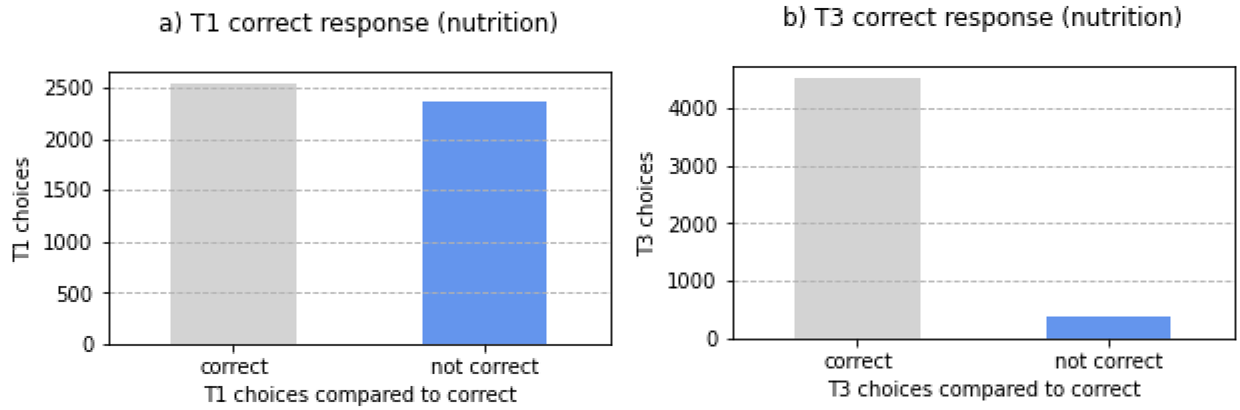

Figure 6.11: Nutrition counts of T1 and T3 CR. a) shows more correct decisions at T1, b) shows more correct decisions at T3, with greater difference than T1.

3.a As T1 given rate increases, amount of CR will increase

It is predicted that as the T1 given information rates increase, the more CR will be made. Figure 6.12 shows an overall increase at T1 and T3 in CR among the rates, except for the largest rate of 0.9. Also, between T1 and T3, T3 has more instances of CR for each rate.

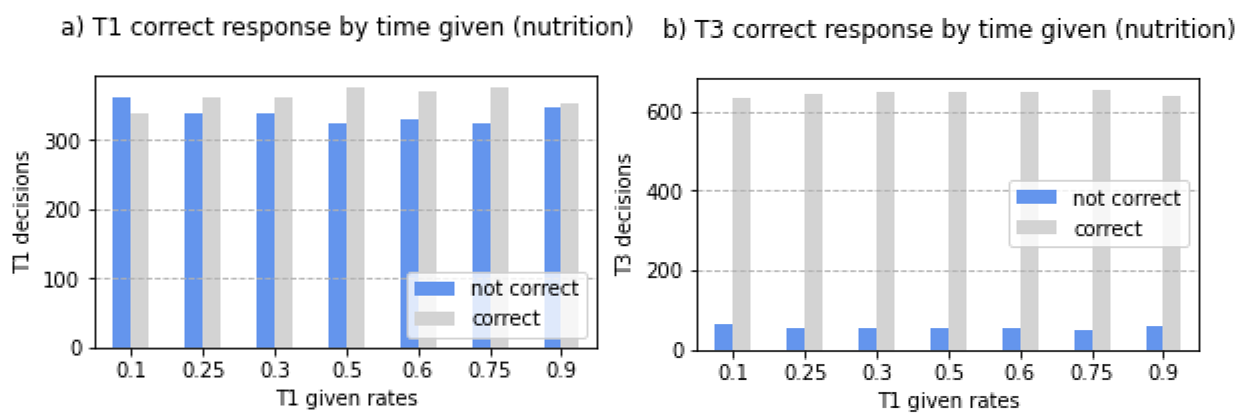

Figure 6.12: Nutrition (3.a) prediction results of a) T1 and b) T3 CR by rate given at T1. There are more CR at T3 than T1 for each time rate.

3.b As T2 given rate increases, the amount of CR will increase

It is predicted that the T2 given rate will reveal the same trends as T1 given rate, where the T2 given rate increases the amount of CR increases. For nutrition, there is not a clear trend for amounts of actual decisions by T2 given rate at T1 or T3 (figure 6.13). Hindsight bias in the nutrition domain is not dependent on T2 given rate as measured by actual decisions.

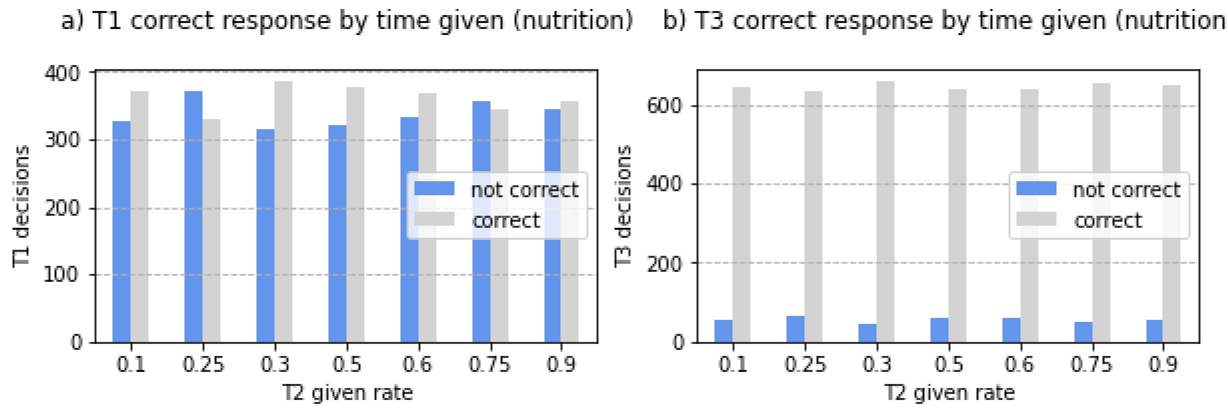

Figure 6.13: Nutrition (3.b) prediction results of a) T1 and b) T3 CR by rate given at T2. There are more CR at T3 than T1.

3.c As time between decisions increases, CR will increase at T3

With time between decisions, it is predicted that as time increases, the more CR are made. Figure 6.14 shows that T1 decisions are independent of time between, with T3 dependent on time between decisions. There are not any incorrect responses after eight weeks between decisions, showing time between decisions to be a strong source of hindsight bias as measured by CR in the nutrition domain.

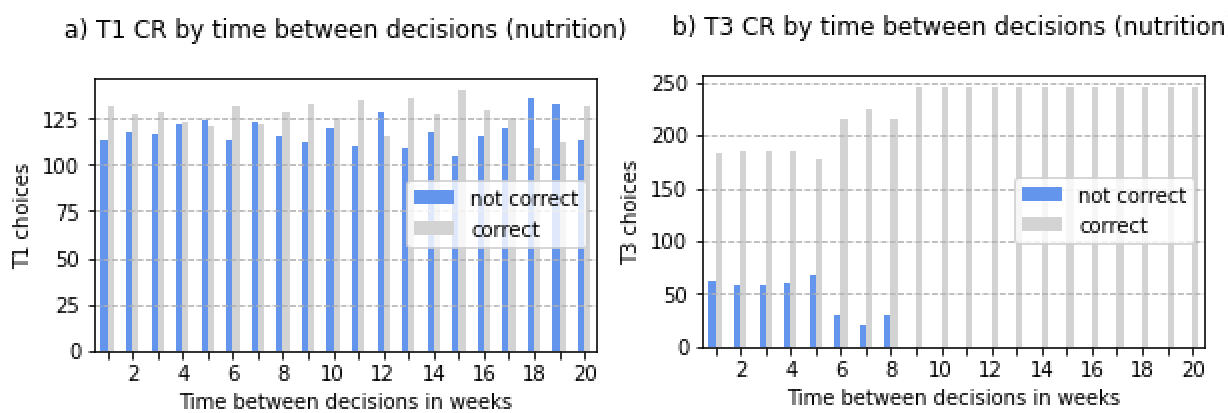

Figure 6.14: Nutrition (3.c) prediction results of a) T1 and b) T3 CR by time between decisions. There are more CR at T3 than T1.

Within the third category of predictions for nutrition, rates given at T1 and T2 are revealed to be positive but weak indicators for hindsight bias as measured by CR. Time between decisions is a strong source of impairment for hindsight bias. Overall, nutrition has a 78.07% increase from T1 to T3 of CR (figure 6.11).

6.3 Combined COVID-19 and Nutrition

This third section of results compares and contrasts the conclusions from the two different domains. Overall, it is predicted that *COVID-19 will reveal greater hindsight bias than nutrition*. Examination of each prediction category demonstrates the nutrition domain to be associated with greater hindsight bias in two of the three categories.

The first prediction category uses the HBC equation to measure hindsight bias. In comparing the averages of HBC for nutrition, cases, and deaths, nutrition results contain higher values for each source of impairment (figures 6.8-11) and more hindsight bias. COVID-19 results show deaths to have higher HBC averages than cases (figures 6.1-3).

By comparing the percent differences in confidence values of T1 and T3, death results are found to demonstrate the most hindsight bias, with cases also being higher than nutrition (cases: 4.65%, deaths: 29.05%, nutrition: 0.028%). This consequence is attributed to the added subjectivity of COVID-19 deaths and cases as compared to nutrition. Similarly, COVID-19 death results are concluded to have more subjectivity than cases.

The third category of predictions measures hindsight bias by comparing the CR of T1 and T3. Nutrition is revealed to have the greatest percent difference of CR, leading nutrition to have the most hindsight bias in this category. The percent difference for nutrition is 78.07%, which is greater than the 29.05% difference of COVID-19 deaths and 4.65% difference of cases.

Through comparison of the results from each prediction category, nutrition is shown to be associated with greater hindsight bias in a majority of the categories. It is concluded that the nutrition domain has more hindsight bias, followed by COVID-19 deaths, and then COVID-19 cases.

CHAPTER 7

DISCUSSION

This section discusses the results in the context of this project's motivations and contributions, the limitations to this project, and directions for future research. Going back to the motivations behind this project, I hoped to create a usable and complete simulation of hindsight bias accounting for two of the three groupings of theoretical cognitive processes underlying hindsight bias. This was accomplished using RAFT and the TTB heuristic with Python in the Anaconda environment. The completion of a computational model in this context depicts the feasibility of expanding computational modeling to research in hindsight bias and psychology. Additionally, applying computational simulation to the specific cognitive process of RAFT shows the necessity of research using different techniques since they highlight different aspects of cognitive processes. Computer simulation allows for a more transparent approach to analysis of decision making. Deeper understanding in regard to the complexity behind cognitive processes, and those related to hindsight bias, can be further explored with the baseline feasibility achieved with this project.

Additionally, I hoped this thesis would provide insight into what a comprehensive model of hindsight bias encompasses, since there is not a single, unified representation in the literature. Testing shows that more often than not hindsight bias occurs as expected, and is thus a satisfactory baseline computational model of hindsight bias. This model works well as a foundation for a unified, comprehensive model, but there are various limitations that need to be addressed before generalization is possible. Analysis of these limitations helps guide the formalization of the requirements needed for a comprehensive model. First it is necessary to compare the computational results to those from human experiments. An experimental comparison will quantify the model's similarity to the biological cognitive processes related to hindsight bias in a way that goes beyond comparison to theoretical predictions that this paper provides. Also, contextualization of the model's results is essential in understanding the degree of adequacy of computational simulation in studying hindsight bias. The second limitation that frames what is needed in a unified model is this model's inattention to the third grouping of possible underlying cognitive processes for hindsight bias, metacognitive motivations. Without expansion to account for motivation as a role in hindsight bias, hindsight bias may not be fully accounted for. One way to expand this research to include motivation would be to introduce a reward function dependent on the model's ability to correctly make a decision. The third way this model comments on creating a unified model of hindsight bias is this simulation's ability to only handle binary decision making tasks. Expansion to handling various types of tasks is necessary for generalization of hindsight bias and should be included in a comprehensive model.

The third general motivation behind this project was to examine whether context plays a role in hindsight bias. Using the model in two different settings, COVID-19 and nutrition, we see that the context of nutrition led to higher levels of hindsight bias. This result contradicts the original prediction that social urgency of the COVID-19 domain would lead to more hindsight bias. But these results do not imply that social urgency was not central in the decision making process for each domain. Instead, the results should be understood in the context of various limitations that may have impacted the performance of the classifiers used in the computational model. First, the target labels for each dataset were derived differently. The COVID-19 labels were a calculation of differences between days, while the nutrition labels were found with the unsupervised K-Means Clustering algorithm. While the clusters were examined, each food instance in the dataset was not manually corrected. Depending on an algorithm to divide the dataset into meaningful clusters could have been one reason for the observed degree of hindsight bias in nutrition. Second, it is important to note limitations in the datasets themselves. The differences in hindsight bias could be a consequence of the sizes of the datasets, where there are many more instances for training and testing in the nutrition scenario. With less data to learn from, it is possible that the model overfit the COVID-19 dataset and resulted in less hindsight bias. Overall, these three kinds of limitations provide a structure in how to expand this project in meaningful ways. l.

The fourth general motivation for this project was to better understand how the underlying cognitive processes of hindsight bias can translate to machine learning. In the field of machine learning, algorithms have recently been exposed to regularly succumb to unintended cue learning, or shortcut learning, from the type and quality of the datasets used (41). This project's simulation of hindsight bias functions as a foundation for analyzing the combination of cognitive processes with various machine learning strategies. The results presented in this paper emphasize the importance of bringing awareness to shortcut learning in hindsight bias and machine learning due to the fact that many of the predictions occur as expected. Since the simple heuristic used in this computational model focuses on the best cues, hindsight bias may be able to be mitigated in serious decision-making tasks in the field of machine learning by controlling for cue values. A future approach would be to further explore cue learning in machine learning by augmenting and manipulating the training data to remove unintended cues.

In summary, this project accomplished its main goals while providing a new computational contribution to the wide literature on hindsight bias. Despite limitations, there are still various ways this research can be revisited and expanded, influencing not just the field of psychology but also machine learning.

BIBLIOGRAPHY

Ash, I. K. (2009). Surprise, memory, and retrospective judgment making: Testing cognitive reconstruction theories of the hindsight bias effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35(4), 916–933.

Ash, I. K., & Wiley, J. (2008). Hindsight bias in insight and mathematical problem-solving: Evidence of different reconstruction mechanisms for metacognitive vs. situational judgments. Memory & Cognition, 36, 822–837,

Bernstein, D.M., Aßfalg, A., Kumar, R., & Ackerman, R. (2016). Looking Backward and Forward on Hindsight Bias.

Bernstein, D. M., Erdfelder, E., Meltzhoff, A. N., Peria, W., & Loftus, G. R. (2011). Hindsight bias from 3 to 95 years of age. Journal of Experimental Psychology: Learning, Memory, and Cognition, 37, 378–391

Blank, H., & Nestler, S. (2007). Cognitive process models of hindsight bias. *Social Cognition, 25*(1), 132–146. https://doi.org/10.1521/soco.2007.25.1.132

Blank, H., Nestler, S., von Collani, G., & Fischer, V. (2008). How many hindsight biases are there? Cognition, 106, 1408–1440.

Breiman, L., Friedman, J., Olshen R., Stone, C. (1984). Classification and Regression Trees. Wadsworth, Belmont, CA, 1984.

Brennen, J. S., Simon, F. M., & Nielsen, R. K. (2021). Beyond (Mis)Representation: Visuals in COVID-19 Misinformation. The International Journal of Press/Politics, 26(1), 277–299.

Calvillo, D. (2012). Working memory and the memory distortion component of hindsight bias. *Memory, 20*, 891 - 898.

Centers for Disease Control and Prevention (CDC). COVID-19 Case and Death Surveillance Public Use Data with Geography. Surveillance Review and Response Group.

Consumer Reports. (2020, 2021). Consumer Reports American Experiences Survey [Dataset]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research.

Coolin, A., Erdfelder, E., Bernstein, D.M. *et al.* (2015). Explaining individual differences in cognitive processes underlying hindsight bias. *Psychon Bull Rev* 22, 328–348. https://doi.org/10.3758/s13423-014-0691-5

Denil, M., Matheson, D., de Freitas, N. (2014). Narrowing the Gap: Random Forests In Theory and In Practice. Proceedings of the 31 st International Conference on Machine Learning. Vol 32.

Erdfelder, E., and Buchner, A.. Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24(2):387, 1998.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. Journal of Experimental Psychology: Human Perception and Performance, 1, 288– 299.

Fischhoff, B., & Beyth, R. (1975). “I knew it would happen”: Remembered probabilities of once future things. Organizational Behavior and Human Performance, 13, 1–16.

Geirhos, R., Jacobsen, JH., Michaelis, C. *et al.* Shortcut learning in deep neural networks. *Nat Mach Intell* 2, 665–673 (2020). https://doi.org/10.1038/s42256-020-00257-z.

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The take the best heuristic. In *Simple heuristics that make us smart* (pp. 75-95). Oxford University Press.
Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. Psychological Review, 98, 506–528.
Gigerenzer, G., & Todd, P. M. (1999). Simple heuristics that make us smart. Oxford University Press.
Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta–analysis of research on hindsight bias. Basic and Applied Social Psychology, 26, 103–117.
Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107*(3), 311–327. https://doi.org/10.1037/0033-2909.107.3.311.
Haytowitz, D., Ahuja, J., Wu, X., Somanchi, M., Nickle, M., Nguyen, Q., Roseland, J., Williams, J., Patterson, K., Li, Y., Pehrsson, P. (2019). USDA National Nutrient Database for Standard Reference, Legacy Release. Nutrient Data Laboratory, Beltsville Human Nutrition Research Center, ARS, USDA.
Hell, W., Gigerenzer, G., Gauggel, S. Mall, M., and Müller, M.. Hindsight bias: An interaction of automatic and motivational factors? Memory & Cognition, 16(6):533–538, 1988.
Hertwig, R., Fanselow, C., & Hoffrage, U. (2003). Hindsight bias: How knowledge and heuristics affect our reconstruction of the past. Memory, 11, 357–377. doi: 10.1080/09658210244000595. PMID: 14562868.
Hoffrage, U., & Pohl, R. (2003). Research on hindsight bias: A rich past, a productive present, and a challenging future. Memory, 11, 329–335.
Hugh, TB, Dekker, SW. (2009). Hindsight bias and outcome bias in the social construction of medical negligence: a review. J Law Med.;16(5):846-57. PMID: 19554863.
Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in science & engineering, 9(3), 90.
International Food Information Council. 2021 Food & Health Survey. 19 May 2021. [https://foodinsight.org/2021-food-health-survey/]
IPSOS. (2020, 2021). Consumer Reports Survey [Dataset]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research.
Jahromi, A.H., and Taheri,M. "A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features," *2017 Artificial Intelligence and Signal Processing Conference (AISP)*, 2017, pp. 209-212, doi: 10.1109/AISP.2017.8324083.
Knoll, M. A. Z., & Arkes, H. R. (2017). The Effects of Expertise on the Hindsight Bias. Journal of Behavioral Decision Making, 30(2), 389–399. https://doi.org/10.1002/bdm.1950.
Louie, T. A. (1999). Decision makers' hindsight bias after receiving favourable and unfavorable feedback. Journal of Applied Psychology, 84, 29–41.
Madarász, K. (2008). Information Projection: Model and Applications. *FEN: Behavioral Finance (Topic).*
Mahdavi, S., & Rahimian, M. (2017). Bias Impedes Learning. Proceedings of Machine Learning Research 58, 111-127.

Meuer, M.W., Nestler, S., & Oeberst, A. (2021). Debiasing Media Articles – Reducing Hindsight Bias in the Production of Written Work. *Journal of applied research in memory and cognition*.
McKinney, W. (2010). Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference (Vol. 445, pp. 51-56).
Oliphant, T. E. (2006). A guide to NumPy (Vol. 1, p. 85). USA: Trelgol Publishing.
Müller, P. and Stahlberg, D. (2007). The role of surprise in hindsight bias: A metacognitive model of reduced and reversed hindsight bias. Social Cognition, 25(1):165.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of\ Machine Learning
Research, 12, 2825–2830
Pezzo, M. (2011). Hindsight bias: A primer for motivational researchers. Social and Personality Psychology Compass, 5(9):665–678.
Pohl, R.F. (2007). Ways to assess hindsight bias. Social Cognition, 25(1) 14–31.
Pohl, R.F., Bender, M., & Lachmann, G. (2002). Hindsight bias around the world. Experimental Psychology, 49, 270–282.
Pohl, R. F., Hardt, O., & Eisenhauer, M. (2000). SARA: Ein kognitives Prozessmodell zur Erklärung von Ankereffekt und Rückschaufehler [SARA: A cognitive process model explaining the anchoring effect and hindsight bias]. Kognitionswissenschaft, 9, 77–92.
Python Software Foundation. Python Language Reference. Website http://www.python.org
Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. Perspectives on Psychological Science, 7, 411– 426. https://journals.sagepub.com/doi/full/10.1177/1745691612454303.
Sanna, L. J., & Schwarz, N. (2007). Metacognitive experiences and hindsight bias: It's not just the thought (content) that counts. Social Cognition, 24(1), 185–202.
Schwarz, S. & Stahlberg, D. (2003). Strength of hindsight bias as a consequence of meta-cognitions. Memory, 11(4-5):395–410.
Taunk, Kashvi & De, Sanjukta & Verma, Srishti & Swetapadma, Aleena. (2019). A Brief Review of Nearest Neighbor Algorithm for Learning and Classification. 1255-1260. 10.1109/ICCS45141.2019.9065747.
2021 Food & Health Survey. 19 May 2021. https://foodinsight.org/2021-food-health-survey/.
Van Rossum, G. (2020). The Python Library Reference. Python Software Foundation.

APPENDIX A

DATA SUMMARIZATION

A.1 COVID-19

From the CDC, data was collected from the COVID-19 tracker dataset. This tracker data is updated daily, beginning January 22, 2020, and is based on records provided by states and territories. Even though this tracker data is being used as a foundation for this project, it does not imply this data is assumed to be completely accurate. There are delays and mistakes in reporting and testing, all those infected do not seek medical care, and there is no way to ensure states and territories record their cases and deaths accurately. Yet this data is utilized because it aggregates information from states and territories, the National Vital Statistics System (NVSS) for death certification, and many local health departments and hospitals. This dataset contains multiple features, such as the submission date, the jurisdiction reporting, that location's total number of cases and total number of deaths, the confirmed cases and deaths, the number of probable new cases and deaths, as well as the number of new cases and deaths for that day. The tracker dataset includes one datetime object, one string feature (the jurisdiction), and the rest of the attributes are represented as integers.

In addition to quantitative information about the cases and deaths relating to COVID-19, it was necessary to also acquire qualitative data, which is why the Consumer Reports and IPSOS surveys are included. In the Consumer Reports survey, which was administered by NORC at the University of Chicago, a nationally representative sample of 2,000 residents responded online and by phone to a variety of questions. This study is multi-mode, but divided into sections, where only relevant COVID-19 items have been preserved. This survey includes questions about COVID-19 concern over the next month and next six months, the vaccine, safety level during multiple activities, and issues with financial assistance. A margin error of +/- 2.93 percentage points was found at a confidence level of 95 percent.

In the weekly tracker survey from IPSOS, consumer's attitudes are measured on various COVID-19 related topics, including personal perceived threat of COVID-19, habits and rituals, media and entertainment, comfort relating to vaccinations and masks, finances and spending, and alcohol consumption, for a total of nineteen questions. Each week, about 1,100 individuals from the continental U.S. complete this survey online. From the targeted sample results, IPSOS analyzes and transforms the results to fit a more general sample. Each survey has a credibility interval of +/- 3.3 percentage points and confidence interval of +/- 4.8 percentage points.

A.2 Nutrition

In the USDA nutrition dataset, all entries have an associated identification number. Additionally, each entry belongs to a given Food Group, and has other non-integer attributes of a short description, a longer description, its common name and its scientific name. The rest of the attributes are integers that explain different levels of nutrition for each entry. For example, some attributes are protein (in g), calcium (in mg), and zinc (in mg). Importantly, not all attributes are on the same scale, i.e. some are measured in grams while others in milligrams.

The qualitative data for the nutrition domain is survey information concerning health, nutrition, and eating behaviors. The survey was conducted by Greenwald Research, using Dynata's consumer panel, formatted as an online survey with 1,014 Americans ages 18 to 80 participating. In order to ensure the results are reflective of the 2020 American population, the results were weighted according to age, education, gender, race/ethnicity and region. The relevant questions and answers are under the Perceptions about Health and Nutrition section of the survey. Two questions and answers are preserved in the context of this project: what kinds of foods are most likely to cause weight gain and what defines a food to be healthy.

APPENDIX B

CLASSIFIERS AND HYPERPARAMETERS

The classifiers used in this paper are K Nearest Neighbors (KNN) Classifier, Decision Tree (DT) Classifier, and Random Forest (RF) Classifier. All classifier hyperparameter tuning was done through GridSearchCV(), which finds the optimal hyperparameters of a model. Optimal means the hyperparameters that result in the best predictions. Before testing, each classifier was trained and tested with a 70-30 split for each dataset. A baseline for each model was determined by training and testing on the default model hyperparameters. After, Grid Search was applied for each classifier in the context of each dataset in order to get the optimal hyperparameters for COVID-19 deaths, COVID-19 cases, and nutrition.

KNN is known for its simplicity and effectiveness, and works by classifying data based on its neighbors using Euclidean distance. Decision Tree classification is a tree based technique that outputs a decision from conditions for each node. DTs are advantageous because they are easy to understand and interpret. Random forests are a type of ensemble method that combines the predictions of several trees that are each trained in isolation. It uses averaging over the subsamples of data to improve predictive accuracy.

The particular hyperparameter used for each dataset are detailed below.

COVID-19 Cases

- KNN – n_neighbors=10, p=1, weights=’distance’
- DT – max_depth=2, criterion=’gini’
- RF – max_features=4, n_estimators=30, bootstrap=True

COVID-19 Deaths

- KNN - n_neighbors=3, p=1, weights=’uniform’
- DT - max_depth=2, criterion=’gini’
- RF - max_features=6, n_estimators=100, bootstrap=True

Nutrition

- KNN - n_neighbors=3, p=1, weights='distance'
- DT - max_depth=12, criterion=’gini’
- RF - max_features=8, n_estimators=100, bootstrap=False

APPENDIX C

ADDITIONAL RESULTS

This section offers general results and statistics from testing that are not defined within the predictions section. First, additional results are given for COVID-19 cases and deaths, then for nutrition.

C.1 COVID-19

C.1.1 Classifiers

Table C.1 shows the number of times each classifier was used for cases and deaths, divided by T1 and T3 decisions. For cases, RF was applied the most at T1 and the most at T3. For deaths, RF was applied the most at T1 and T3. These results show that the ensemble method was most frequently chosen in both COVID-19 settings.

Table C.1: Classifiers utilized at T1 and T3 decisions for COVID-19 cases and deaths.

		Classifier counts		
		KNN	RF	DT
Cases	T1	1520	**2160**	480
	T3	168	**1545**	361
Deaths	T1	0	**2640**	1520
	T3	179	**1558**	351

Table C.2 shows the accuracies of each classifier, for T1 and T3. For both cases and deaths, there is an increase in accuracy, precision, recall, and F1 score from T1 to T3. The accuracy difference for cases is larger than for deaths, meaning cases should have higher levels of hindsight bias. This is supported by the evidence in section 1 of the results chapter.

Table C.2: Classifier accuracies at T1 and T3 for COVID-19 cases and deaths.

			Accuracy	Precision	Recall	F1
Cases	T1	mean	0.9886	0.9940	0.9876	0.9897
		std	0.0373	0.0299	0.0539	0.0348
	T3	mean	0.9975	0.9984	0.9964	0.9972
		std	0.0109	0.0107	0.0207	0.0123
Deaths	T1	mean	0.9908	0.9904	0.9923	0.9904
		std	0.0291	0.0414	0.0385	0.0308
	T3	mean	0.9969	0.9980	0.9958	0.9967
		std	0.0130	0.0141	0.0215	0.0136

C.1.2 Decisions made at T1 and T3

As explained in the predictions section, hindsight bias is shown when decisions change. Here the amount of decisions (either increase or decrease) are given at T1 and T3 in figure C.1. For cases, there are more decisions of increase at T3 than at T1. Deaths show the same trend as cases between T1 and T3.

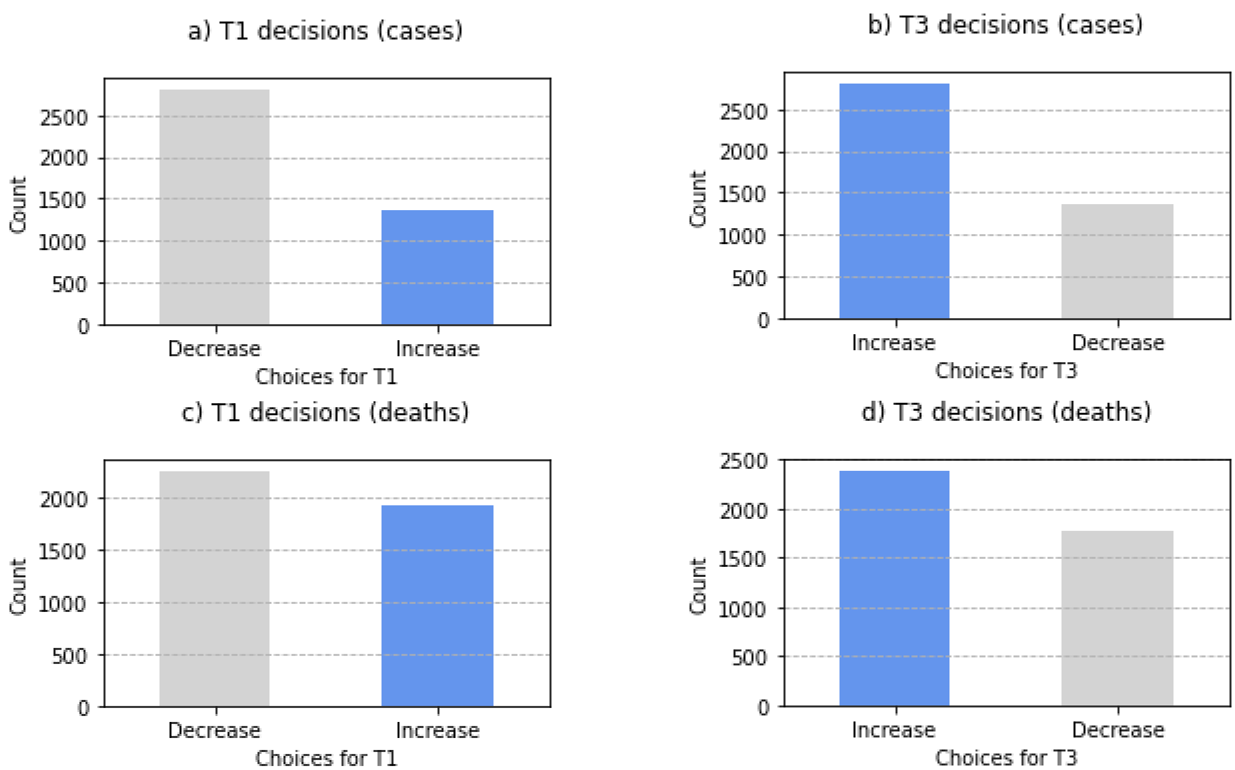

Figure C.1: Decisions made at T1 and T3 for COVID-19. a) shows more instances of decrease for cases at T1, b) shows more instances of increase for cases at T3, c) and d) show the same for deaths as cases.

C.1.3 Similar decisions between T1 and T3

In comparing similar decisions, there are more instances of similar decisions between T1 and T3 for cases than for deaths, which is supported by the evidence in the results chapter.

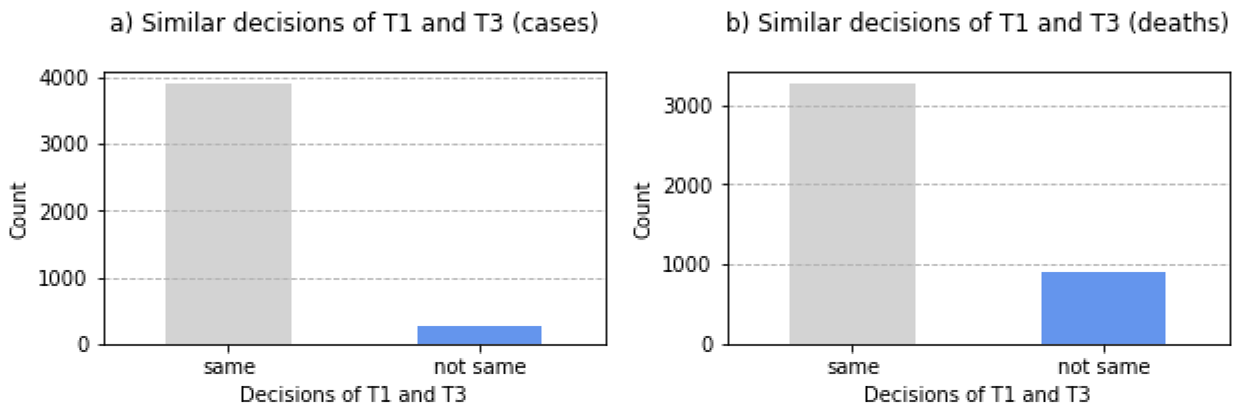

Figure C.2: Similar decisions between T1 and T3 for COVID-19. a) shows the results for cases, b) shows the results for deaths, with more different decisions than cases.

C.1.4 Recall and Reconstruction value counts

Since recall and reconstruction were determined randomly, not dependent on any source of impairment, the number of times for each deliberation method should be about equal. The counts of each in figure C.3 show this to occur.

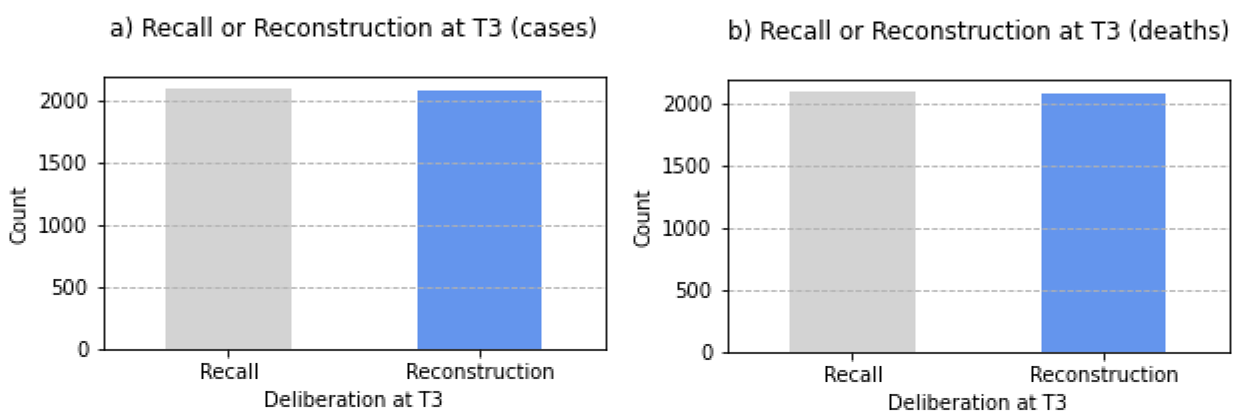

Figure C.3: Recall or Reconstruction for COVID-19 a) cases and b) deaths.

C.1.5 HBC levels

From the HBC equation, each test instance was grouped into either high or low categories of HBC, according to distribution in each scenario. The threshold for cases is 0.194 and is 0.199 for deaths. There are more instances of high HBC for deaths than for cases.

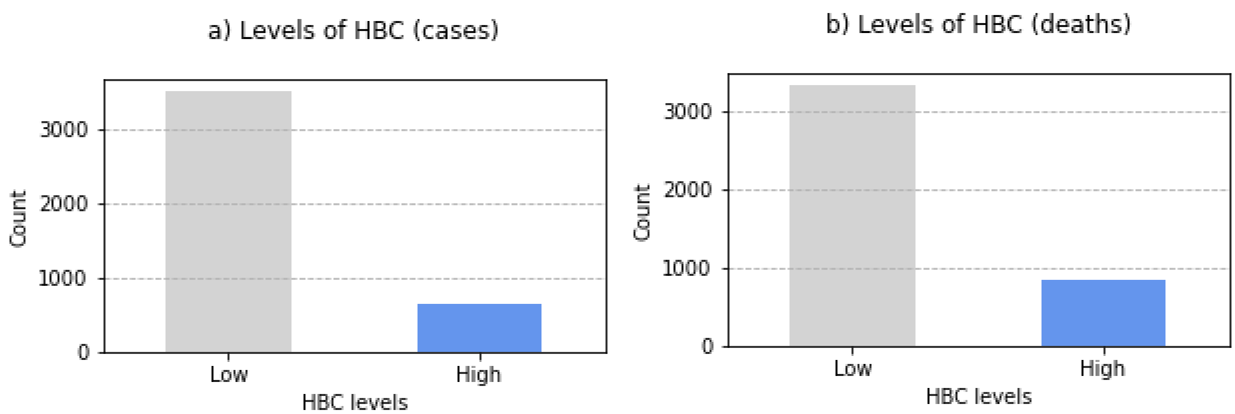

Figure C.4: HBC levels between T1 and T3 of COVID-19 a) cases and b) deaths, with more high changes than cases.

C.2 Nutrition

C.2.1 Classifiers

Table C.3 shows the number of times each classifier was used in the nutrition results, divided by T1 and T3 decisions. These results show that the ensemble method performed the best so was most frequently chosen in both COVID-19 settings.

Table C.3: Classifiers utilized at T1 and T3 decisions for nutrition.

		Classifier counts		
		KNN	**RF**	**DT**
Nutrition	**T1**	0	**3774**	1126
	T3	0	**1944**	506

Table C.4 shows the accuracies of each classifier, for both T1 and T3. Nutrition results show accuracy, precision, recall and F1 score increase from T1 to T3. This is supported by the evidence in section 1 of the results chapter.

Table C.4: Classifier accuracies at T1 and T3 for nutrition.

			Accuracy	Precision	Recall	F1
Nutrition	T1	mean	0.9973	0.9985	0.9969	0.9977
		std	0.0028	0.0027	0.0036	0.0024
	T3	mean	0.9985	0.9980	0.9981	0.9987
		std	0.0012	0.0141	0.0017	0.0011

C.2.2 Decisions made at T1 and T3

As explained in the predictions section, hindsight bias is shown when decisions change. Here the amount of decisions (either healthy or unhealthy) are given at T1 and T3 in figure C.5. For nutrition, there are only 70 more instances of healthy chosen at T3 than T1. Since there is not a significant difference in amounts of decisions but we know nutrition is associated with more hindsight bias than COVID-19, these results are attributed to the ratio of decisions right at T1, but made on the wrong instances.

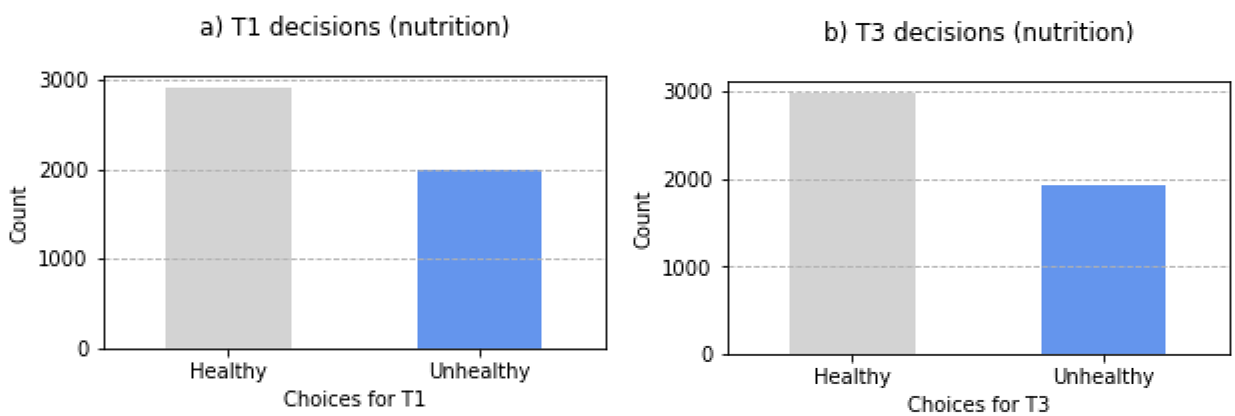

Figure C.5: Decisions made at T1 and T3 for nutrition. a) shows more instances of healthy decisions at T1, b) shows more instances of healthy decisions at T3, only slightly increasing from T1.

C.2.3 Similar decisions between T1 and T3

In comparing similar decisions, there are more instances of similar decisions than different between T1 and T3 for nutrition.

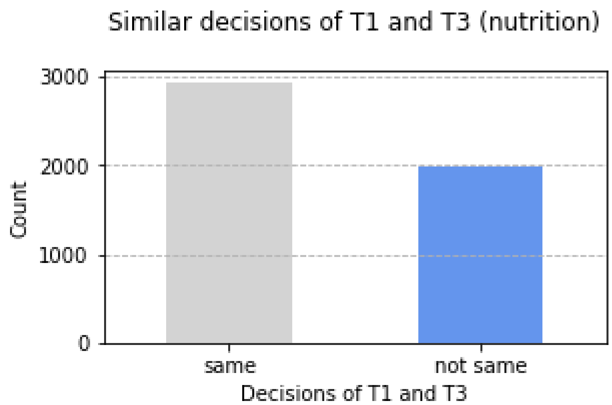

Figure C.6: Similar decisions between T1 and T3 for nutrition, with more same decisions than different.

C.2.4 Recall and Reconstruction value counts

Since recall and reconstruction were determined randomly, not dependent on any source of impairment, the number of times for each deliberation method should be about equal. The counts of each in figure C.7 show this to occur.

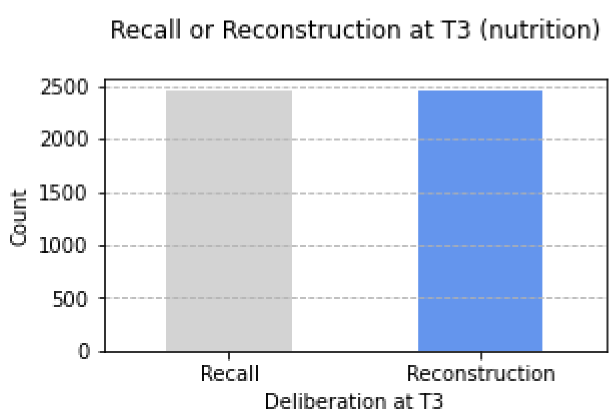

Figure C.7: Recall or Reconstruction for nutrition.

C.2.5 HBC levels

From the HBC equation, each test instance was grouped into either high or low categories of HBC, according to its distribution. The threshold for nutrition is 5.68. There are more instances of low change than high.

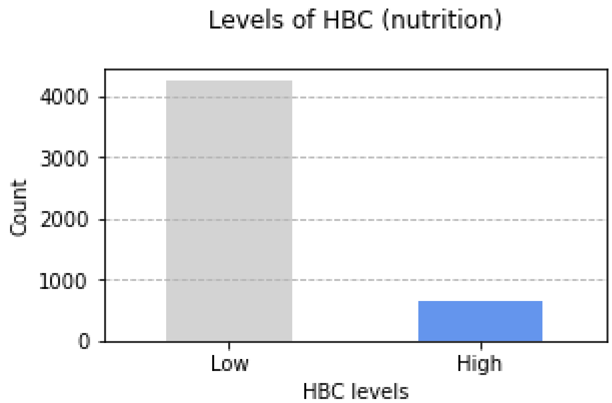

Figure C.8: HBC levels between T1 and T3 of nutrition.