Mechanisms of Optimistically Biased Belief Formation

By Christopher Melamed
Advised by: Laurie Santos and Matthew Jordan

Submitted to the faculty of the Department of Cognitive Science in partial fulfillment of the requirements for the degree of Bachelors of Science

Abstract

Humans change their beliefs in response to new information, but do they do so rationally? Research shows that our belief updating is systematically biased. When individuals receive information that is favorable with respect to their goals, they update their beliefs using that information rationally. When individuals receive unfavorable information, however, they consistently fail to incorporate this information when updating beliefs. Are there mechanisms that underlie this biased belief updating, and, if so, what are they? Our hypothesis is that this phenomenon is the result of something more than simply being wrong or ignoring information. Hints from psychological literature suggest that information, regardless of valence, is encoded within the mind, but isn't necessarily used to update beliefs. The present paper uses cognitive load and lexical decision tasks to examine asymmetric belief updating in individuals who have access to the information needed to form accurate beliefs. Our results indicate that high cognitive load following an individual's exposure to unfavorable information selectively eliminates asymmetry in belief updating. The results of the lexical decision task experiment indicate that asymmetric belief updating is mediated by a process that generally inhibits unfavorable information. Together, these results provide indication that an active, generally inhibitory mechanism is mediating the failure to integrate unfavorable information with novel beliefs at the moment of information retrieval.

Introduction

We all have beliefs about the world. Some of our beliefs are accurate, some are not, but we generally want the things we believe to be true. When faced with evidence that our beliefs are inaccurate, we should want to modify our beliefs to maintain as accurate a representation of reality as possible. As our lives progress, these beliefs should shift slowly toward greater and greater accuracy over time.

Unfortunately, there are many cases in which it seems that beliefs are not merely noisy representations of the truth, but rather they are *systematically biased*. We have a tendency to assess ourselves as above average with respect to virtues like intelligence, attractiveness, honesty, and modesty (Hoorens, 1995). 80% of drivers consider themselves above average drivers, 90% of professors consider themselves above average at teaching, and 87% of MBA students consider themselves above average students (McCormick et al., 1986; Cross, 1977; Zuckerman et al., 2001). Undoubtedly, these findings can be accounted for in part by the human tendency to overestimate one's own abilities compared to others, otherwise known as the above average effect. This phenomenon, perhaps, is the result of these individuals simply not having access to the information necessary to make accurate judgments about themselves. That being said, what about the times when one *does* have access to this information? There are other cases when people are very specifically given the information that is relevant to updating their beliefs and they still fail to do so. Indeed, people hold these optimistic beliefs about their abilities despite access to information like bad grades or having been in car accidents that should indicate otherwise.

This paper is concerned with exploring the mechanisms underlying this failure to incorporate unfavorable information into updated beliefs. First, we will present literature that establishes a tendency to asymmetrically update beliefs. Next, we will discuss potential accounts for why this bias exists and theories about how it might function. We will discuss work that shows that unfavorable information is retained but fails to be used to update beliefs. We then provide evidence for asymmetric belief updating being a result of motivated cognitive processing. Finally, we will present the experimental paradigms we plan to use to explore the cognitive mechanisms underlying asymmetric belief updating.

Work studying human tendencies towards irrationality has uncovered a systematic asymmetry in belief updating (Sharot, 2007). Logically, when presented with novel information, one should update one's beliefs in accordance with Bayes' rule. Bayes' rule states that expectations about quality X shift as a function of information relevant to quality X. For instance, if someone were to ask me how likely I am to experience food poisoning in my life I would guess 50%. This initial estimate is known as my *prior*. Now, imagine I am told that, in actuality, I have a 39% chance of experiencing food poisoning in my life. This actual statistical likelihood of an event's occurrence in the population is referred to as an event's *base-rate.* Now, according to Bayesian updating, if I were to give an updated estimate of my likelihood of experiencing food poisoning following exposure to the event's base-rate, this posterior estimate would incorporate the base-rate into a new, more accurate final estimate. This final estimate is known as the *posterior.* Now, if my beliefs were updated perfectly rationally, my posterior would fall somewhere between 50% and 39%, say, at 43%. This should seem almost painfully obvious, because it is a description of the most logical method of belief updating; a method that allows the individual to modify their beliefs as more information becomes available to them. Almost as obvious should be the fact that this description of belief updating is unrealistically optimistic about human rationality. In practice, work in this area has shown that people are conservative when updating their beliefs regardless of the information that they received (Mobius, 2012). That is, whether the information they receive is positive or negative, individuals update their beliefs less than they rationally should. We are never perfect when it comes to updating our beliefs, but we display certain interesting tendencies.

In addition to being conservative in their updating, people show other biases in how they update their beliefs. Beliefs systematically fail to track with the information available to the belief holders (Eli & Rao, 2012). People update in a way that, though noisy, tracks with Bayes' rule for positive information but not for negative information. To continue from the previous example, having estimated my likelihood of experiencing food poisoning at 50%, if provided with a base-rate of 39%, this would constitute favorable news. This is favorable because experiencing food poisoning is universally recognized as an aversive event, and any information indicating that it is less likely to occur is desirable. In good news scenarios like this, individuals roughly adhere to Bayesian updating and would provide a posterior that falls between their prior and the base-rate (again, say, 43%). If, however, my prior had been 20%, then the base-rate of 39% would be unfavorable news. In bad news scenarios like this, individuals update beliefs more conservatively and in a way that does not reflect Bayesian updating (Eli & Rao, 2012). Here, my posterior might be 23%, remaining relatively unaffected by the base-rate information. While Bayesian updating is how individuals should logically form new beliefs, people tend to update asymmetrically when given new information, largely conforming to Bayes' rule when they get good news, but updating very conservatively when they get bad news.

How can we account for this asymmetric belief updating? Cognitive neuroscientists (Sharot, 2007, 2012) and economists (Eli & Rao, 2012; Mobius, 2012) have replicated this asymmetry in belief updating, but have not answered the question of how people can receive information and not use it. There are many ways one can come to have biased beliefs about the world, but we are focusing on situations in which an individual has all of the information necessary to make accurate estimations yet still fails to do so. Given this stipulation, we have two hypotheses about how this might happen: The first suggests that while people are given relevant information, this information is somehow barred from entering the mind and is therefore not used in belief updating. The second is that the relevant information is encoded, but there is a cognitive mechanism preventing it from being integrated into new beliefs.

If our first hypothesis were true, you would expect people to be unable to remember base-rates that constituted bad news. Recent work shows, however, that unfavorable information is encoded and stored in memory (Sharot, 2011). In one experiment, the researchers asked subjects to estimate the likelihood of experiencing various adverse events. After reporting their priors, the participants received the base-rates for each of these events. Following the base-rates, subjects gave their posterior estimations of the likelihood of each event. When the base-rates were unfavorable, participants displayed a selective failure to update their beliefs. Crucially, in the final section of the experiment, subjects were asked to recall the base-rates provided to them. The results showed that, regardless of the valence of the base-rates given the subjects' priors, participants were able to recall the base-rates with high accuracy (Sharot, 2012). The subjects' consistent ability to remember the base-rates indicates that while the unfavorable information wasn't used to update participants' beliefs, it was encoded and remained accessible in the subjects' memories. This supports our second hypothesis, that there is a cognitive mechanism preventing relevant and available information from integrating with one's new beliefs.

Why would one fail to use information that is available to them? From what we've seen, people are willing to update their beliefs in the presence of good news but not bad news. By definition, good news is attractive and bad news is aversive. Work studying motivated reasoning has established the fact that people have goals, both implicit and explicit, that affect the way they process information and perform tasks (Kunda, 1990). For the most part, these goals lead us to seek out positive, goal affirming information (Stajkovic, 2006). Research on asymmetric belief updating, whether done by cognitive neuroscientists or economists, consistently uses self-relevant tasks in order to evoke these goal-related effects: sometimes these self-relevant tasks are related to the participant's intelligence, while other times these self-relevant tasks are related to life outcomes and well-being (Eli & Rao, 2012; Sharot, 2011; Mobius, 2012). This work also shows that asymmetric belief updating is less pronounced in tasks when there is no ego at stake (Mobius, 2012). Essentially, the information must be self-relevant or there is no reason to asymmetrically update beliefs. If this is the case, then, a motivated cognitive mechanism is selectively ignoring unfavorable information in order to fulfill latent goals (goals to be seen as smart, intelligent, humble, interesting, etc.). This implies that asymmetric belief updating is supported by a motivated, self-serving cognitive process.

Now that we have established what is required for asymmetric belief updating to be possible, the question that remains is: how exactly does asymmetric belief updating work? Recall that, although unfavorable information is encoded, such information is less likely to be fully incorporated into beliefs. What psychological mechanisms would support such a failure to be rational? One likely possibility is that such information is kept out of individuals' updated beliefs through inhibition. Indeed, some work has suggested that inhibition plays a role in asymmetric updating (Sharot et al., 2012). Sharot and colleagues (2012) explored mechanisms underlying asymmetric belief updating using transcranial magnetic stimulation (TMS) to disrupt the region of the brain known as the inferior frontal gyrus (IFG), an area that has been implicated in the updating of beliefs as well as general inhibition (Cools, 2002; Sharot, 2012). Under the influence of TMS, the researchers asked subjects to assess the likelihood of experiencing a set of negative life events. They then provided subjects with the given events' base-rates and asked subjects to report their posteriors. The researchers split the subjects into three groups, those that received TMS to the right IFG, those that received TMS to the left IFG, and a control to which no TMS was applied. When the base-rates that the subjects received were unfavorable with respect to their priors, subjects failed to incorporate the base-rates into their posteriors. In other words, both the control and the right IFG TMS groups displayed asymmetric belief updating. When subjects who received TMS to the left IFG completed this same task, however, they incorporated the base-rates they received into their posteriors regardless of the base-rates' valence given the subjects' priors. That is, TMS disruption of left IFG induced *symmetric* belief updating for both good news and bad news. Neither the right IFG group nor the control group updated symmetrically (Sharot, 2012). Keeping in mind that the left IFG has been implicated in inhibition, this study suggests that inhibition plays a role in asymmetric belief updating.

It is important to note, however, that TMS is a blunt tool for studying cognitive processes for two reasons: localization and temporal precision. It is very difficult to precisely control the extent to which TMS stimulation spreads throughout the brain. Precise regions of the brain often play various roles in various systems and mechanisms; without substantial spatial acuity it is difficult to infer anything from the aforementioned study other than a potential interaction with inhibitory processes. In addition, the duration of the effects of TMS is not controllable. This is an impediment to an experiment like this, in which it is valuable to understand when, throughout the various tasks presented to the subjects, inhibitory processes are acting. The left IFG was disrupted throughout the duration of the study, so while we can infer that the left IFG is somehow involved in asymmetric updating, we cannot make any inferences about how or when it impacts this behavior. While this work (Sharot, 2012) was an important first step in understanding the mechanisms underlying asymmetric updating, it remains unclear how and when inhibition is relevant.

Thus far, we have established that people tend to update beliefs asymmetrically even though all of the information needed to make rational inferences has been received and retained. We have shown that asymmetric belief updating is likely the result of some inhibitory mechanism that acts on behalf of an individual's implicit or explicit goals. Now, our goal is to examine asymmetric belief updating such that we get a better sense of how it occurs and what the role of inhibition might be. This paper presents two experiments intended to explore how this inhibitory mechanism might act.

Cognitive Load

We employed a cognitive load task in order to examine both whether asymmetric belief updating is mediated by an active cognitive process and whether it occurs at the point of information encoding or retrieval. We predict that, if asymmetric belief updating is the result of an active process, increased cognitive load should disrupt this process and subjects should update beliefs symmetrically. Additionally, we placed subjects under load specifically while retrieving their posteriors. If, as we predict, asymmetric belief updating occurs at the moment of information retrieval, then we should see a reduction in asymmetric belief updating as a result of the cognitive load task.

Cognitive load theory is based on the idea that the brain has a finite amount of working processing power. The term refers broadly to a construct that encompasses mental load, mental effort, and performance when an individual is presented with a task (Paas et al., 2003). As the cognitive load placed on an individual increases, lower priority cognitive processes are attenuated. Early studies focused on recognizing and working within the cognitive limitations of the brain in order to create more effective problem solving environments for individuals (Sweller, 1988). Research into lie detection has shown that lying— a case in which people need to hold two pieces of conflicting information at once— is a cognitively taxing mental exercise (Vrij et al., 2006; Vrij et al., 2008). We hypothesized that the same would be true when one is holding two conflicting pieces of information at once but only using one to update their beliefs.

The present study attempted to disrupt asymmetric belief updating via the addition of cognitive load during key moments of information integration. We began by asking subjects to estimate the likelihood of 40 negative life events. Following the presentation of each, we gave the subjects the base-rates for those events—the actual likelihood of the events occurring in the lives of an individual coming from a similar background. We then asked the subject to remember a string of numbers before asking for their posterior estimate of the likelihood of a negative life event. Finally, we asked the subject to recall the numbers they had memorized. We presented each of the 40 events again, each while the subjects were under the cognitive load of this memory task. The memory task is intended to consume the subject's mental processing power and consequently attenuate asymmetric belief updating. Temporally, we presented the memory task here so that the subjects were under cognitive load while giving their posteriors. If asymmetric belief updating requires active mental processing power and occurs when one is reporting one's posteriors, we predict that the cognitive load task should disrupt asymmetric belief updating. If our hypothesis is incorrect and this mechanism functions differently or at a different time, then asymmetric belief updating should remain unaffected by the cognitive load task.

Cognitive Load Experiment Methods

Participants. Forty undergraduate students were recruited from the student subject pool (age range: 19 – 22 years). All subjects gave written informed consent and were given class credit for participation. The study was approved by Yale's Human Subjects Committee.

Stimuli. We presented subjects with forty prompts containing descriptions of negative life events (See: Appendix 1) in a random order. We then provided participants with information about the population-level exposure to each event—these base-rates were based on the likelihood that a young adult living in a western democracy would experience the event. We excluded extremely common and uncommon events from the study such that all events had probabilities between 20% and 70%. The base-rates of the pool of events had a mean and median of 50%.

Procedure

Priors. A computer program presented participants with each of 40 adverse life events in a random order and asked to give their estimates of the likelihood from 0% to 100% that the presented event would happen to them. Participants made estimates by dragging a sliding bar to the desired number. After providing their *prior* for each item, participants received the base-rate for that item, which was presented for 4s. Participants alternated giving prior estimates and receiving base-rate information until they had completed all 40 items.

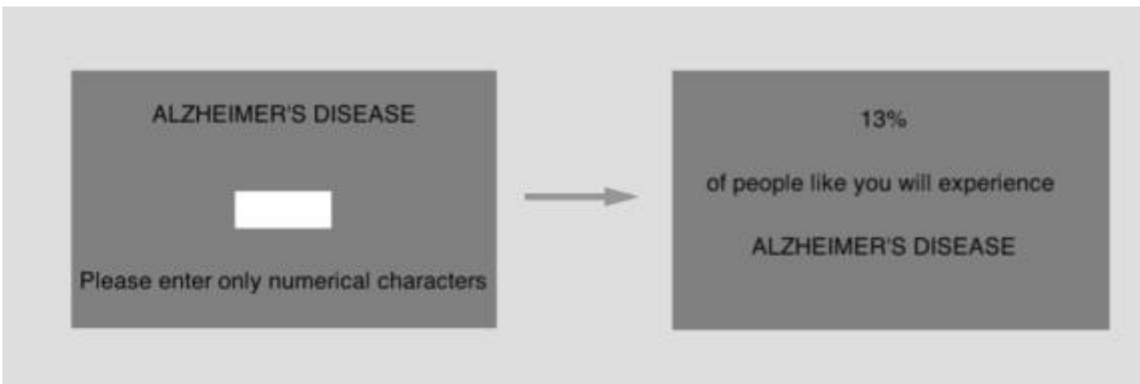

Figure 1: Each trial presented subjects with one of 40 adverse life events and asked for a numerical estimation of the given event's likelihood. Following this estimation, we provided subjects with the base-rate likelihood of the given event.

*Posteriors***.** A prompt told participants that they were about to see the same 40 items again so they could provide beliefs that may have changed in light of the information they received during the priors phase. However, before each event appeared on-screen participants were asked to memorize a string of eight digits. For those in the *high load condition*, these were eight random digits (e.g. 39482754), while those in the *low load condition* saw a single digit repeated eight times (e.g. 33333333). Once participants had memorized the eight digits, they were prompted with an adverse life event and provided their updated posterior. Following the submission of their posterior estimation, we asked subjects to enter the string of digits that they were asked to remember. This pattern then continued; participants again memorized an eight-digit number, were prompted with an event, provided a posterior, and entered the digit they memorized repeated until participants had responded to each of the 40 items.

Memory Test. In the final part of the study, participants had to recall the base-rate information they had received in the priors phase of the study. We asked that they give their best recollection of the accurate base rate they were told earlier.

Coding. For each item, participants could receive *good news* (their prior for this negative event was higher than the base-rate and thus the likelihood of this bad event was less bad than they thought) or *bad news* (their prior was lower than the base-rate and thus the likelihood of this bad event was worse than they thought). Furthermore, for each item, participants could update their beliefs. For simplicity, we computed a fraction-updated value for each response with |prior – posterior| in the numerator and |prior – base-rate| in the denominator. That is, we divided the amount they changed their belief by the amount they could have changed their belief. Past work (Sharot, 2012) has found that participants update more for good news items than bad news items. We should observe such asymmetric updating among participants in the low load conditions. However, if asymmetric updating requires effort during retrieval, we should find that participants in the high load condition update similarly for good news and bad news items.

Results. We predicted that cognitive load would disrupt asymmetric belief updating by attenuating the cognitive processing resources necessary to prevent unwanted information from incorporating with new beliefs. Our results showed that participants displayed asymmetric belief updating under both low load and no load. Participants who were under high cognitive load, however, showed symmetrical belief updating, suggesting a systematic disruption of the mechanism of optimism bias by cognitive load. The fraction participants updated for good news and bad news items, under low load, was 78% and 24%, respectively. Under high load, the fraction participants updated for good news and bad news items was 94% and 62%, respectively. This indicates a near total amelioration of asymmetric belief updating under cognitive load. See *figure 2* below.

We found no main effects of type of news, β = 0.320, CI [-0.071, 0.711], p = 0.109871, or load, β = 0.045, CI [-0.324, 0.414], p = 0.811, on fraction updated. However, we did find evidence of a type of news by load interaction, β = -0.858, CI [-1.604, -0.113], p = 0.024, such that participants in the low load conditions updated more for good news (M = 0.779, CI [0.348, 1.210]) than bad news (M = 0.241, CI [-0.157, 0.639]), β = -0.534, CI [0.049, 1.018], p = 0.031, but participants in the high load condition updated similarly for good (M = 0.937, CI [0.612, 1.262]) and bad news (M = 0.617, CI [0.234, 1.000]), β = -0.333, CI [-0.864, 0.198], p = 0.220.

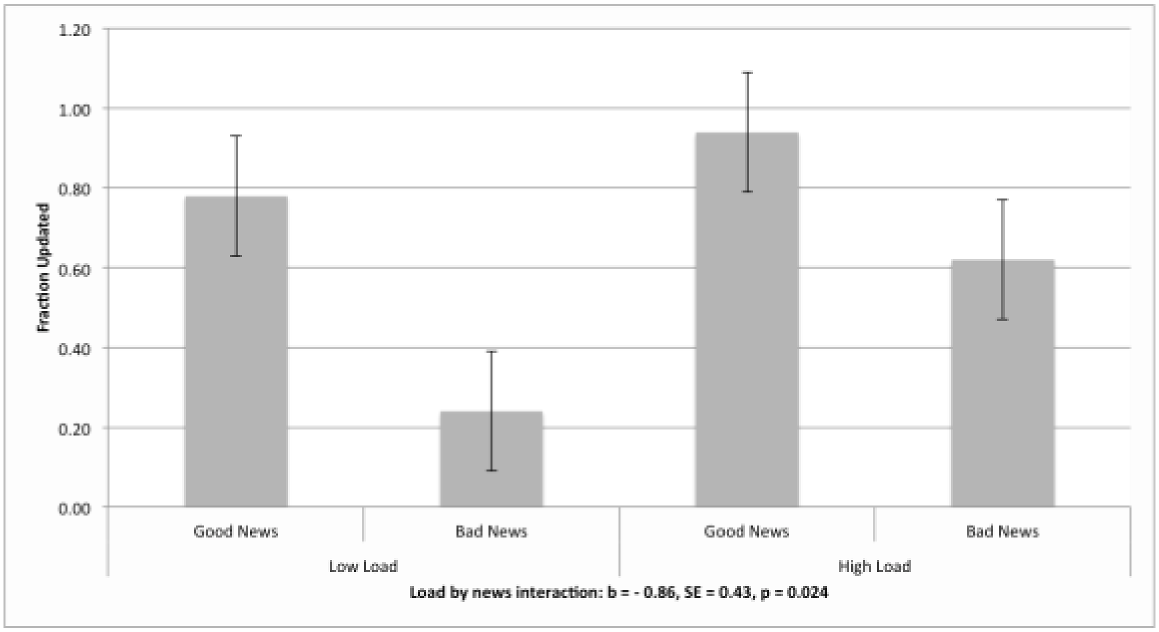

Figure 2: Fraction that beliefs were updated for good news and bad news items organized by low load and high load.

Cognitive Load Experiment Discussion. This study was intended to explore both whether asymmetric belief updating is the result of an active cognitive process and whether it occurs at the moment of encoding or retrieval of novel information. As we hypothesized, our results showed that high cognitive load, as opposed to low cognitive load or no cognitive load, eliminated asymmetric belief updating. The absence of asymmetric belief updating under cognitive load indicates that the mechanism underlying asymmetries in belief formation require cognitive resources and can be attenuated by the presence of tasks that compete for these resources.

The results provide insight in the question of when cognitive resources are needed for asymmetric belief updating. We presented cognitive load tasks specifically when the subjects were retrieving information in order to report their posteriors. Subjecting participants to load only during the estimation of their posteriors gives us valuable insight into when asymmetric belief updating is occurring. If asymmetric belief updating occurred as a result of processing at the moment of encoding information, then cognitive load at the moment of retrieval should have had no effect. These results thus indicate that unfavorable information is inhibited from integrating with new beliefs during retrieval specifically.

One question that remains is whether failure to integrate unfavorable information with new beliefs is the result of general inhibition or rather specific inhibition of the integration of unfavorable information with belief formation. In an attempt to answer this question, we ran a second study employing a lexical decision task. We hypothesized that, if asymmetric belief updating is mediated by general inhibition, then subjects would show increased latency if asked to complete a task that requires the processing of this information.

Lexical Decision Making

The TMS study described earlier indicated that inhibition may play a role in peoples' asymmetric updating of beliefs. If inhibition plays a role in asymmetric updating, it might be the case that participants are actively inhibiting the thought of events for which they have received bad news. Such a general inhibition could result in increased latency when asked to perform tasks requiring the processing of this information. In order to test for this effect, we plan to employ a *lexical decision task* that requires subjects to identify stimuli relating to both favorable and unfavorable information as quickly as they can. According to our hypothesis, we expect that accuracy and response times should be worse for words related to unfavorable information as a result of general inhibition.

The lexical decision task is an experimental technique that measures the speed with which subjects identify strings of letters as words or non-words. The task was developed as a means for studying the mechanisms of semantic memory, particularly as it pertains to language and word recognition (Meyer & Schvaneveldt, 1971). Often used to study priming effects, the lexical decision task shows that subjects are faster when presented with words that are in similar semantic categories to those that have already been presented (Balota & Lorch, 1986).

Once again, we asked to estimate the likelihood of 40 negative life events. After presenting each event we gave the base-rate likelihood of that event occurring in the lives of an individuals coming from similar backgrounds to the subjects. At this point we asked subjects to identify strings of letters presented onscreen for brief intervals as either words or non-words. We hypothesized that participants would be slower to recognize bad news items as words than they would be to recognize good news items as words. This result would suggest that asymmetric belief updating is a result of general inhibition of thoughts pertaining to bad news items.

Lexical Decision Experiment Materials and Methods

Participants. Forty undergraduate students were recruited from the student subject pool (age range: 19 – 22 years). All subjects gave written informed consent and were given class credit for participation. The study was approved by Yale's Human Subjects Committee.

Stimuli. This study's stimuli are the same as in study 1.

Procedure.

Priors. This study's methods for gathering the subjects' priors are the same as in study 1.

Lexical Decision Task. The lexical decision task measures subjects' response times when presented with strings of letters and asked to assess them as words or non-words. The task is typically used to measure priming effects on certain words. Here we tracked accuracy and response times in order to gain insights into rates of inhibition of various stimuli. We presented participants with strings of letters written in green or red text for a period of 4s. All words, no matter the length or color, had appeared in the 40 adverse life events presented earlier. Any word under 4 letters long appeared in green and was always correctly spelled. Words in red were either correctly spelled or randomly jumbled such that they no longer spelled an English word. Following each stimulus, the screen flashed a sting of '#'s so as to obstruct afterimages of the presented strings. We asked participants to decide between ‘word’ and ‘non word’ for the presented strings (green strings were always words but the subjects still had to submit a response) We recorded the response times and accuracy of each of the subjects’ decisions. This task examines the relationship between priming and inhibition. We predicted that the response times and accuracy of the subjects’ responses would vary as a function of whether the presented stimuli were primed by event information that was either positive or negative.

Memory Test. In the final part of the study, we asked participants to recall the base-rate information they had received in the *priors* phase of the study. We asked that they give their best recollection of the base-rates they were told earlier.

Results. We predicted that if asymmetric belief updating results from general inhibition, then subjects should inhibit the thought of words relating to unfavorable information. As a result, we would expect that word/non-word discrimination accuracy would be lower and reaction times would be longer when subjects were assessing the word/non-word status of stimuli related to events whose base-rates were unfavorable given the subject’s priors. On the other hand, if inhibition is specific to integrating the base-rates with beliefs, there shouldn’t be a difference in the response times or accuracy identifying good news vs. bad news items as words. Our data indicated that, for stimuli derived from good news items, response times were faster for stimuli that were judged to be words rather than non-words. On the other hand, for stimuli derived from bad news items, response times were the same for stimuli that were judged to be words and non-words. We found no correlation between the accuracy of the subjects' assessments and whether the stimuli were derived from good news or bad news items.

Because the response times were non-normally distributed, we conducted all analyses using an inverse transformation. We found a main effect of whether a word or non-word was presented, $\beta = 0.032$, CI [0.004, 0.061], $p = 0.025$, such that participants were faster to response on trials in which a word was presented. We found no evidence of a main effect of type of news, $\beta = 0.001$, CI [-0.028, 0.031], $p = 0.921$. However, the main effect of the presence of a word was qualified by a word presence by news type interaction, $\beta = 0.074$, CI [0.017, 0.130], $p = 0.010$, such that among good news trials, $\beta = 0.071$, CI [0.030, 0.113], $p = 0.001$, we observed faster responses when words were present, whereas among bad news trials, $\beta = -0.042$, CI [-0.091, 0.008], $p = 0.098$, participants were marginally slower to respond to words than non-words.

Lexical Decision Experiment Discussion. This study was intended to explore the role of inhibition in asymmetric belief updating. Specifically, we hoped to elucidate whether asymmetric belief updating is a result of general inhibition of unfavorable information or specific inhibition of the integration of unfavorable event base-rates with newly formed beliefs. We predicted that subjects would display slower response times to stimuli derived from bad news items than those derived from good news items. Our results showed that when subjects were presented with stimuli derived from good news items, subjects were faster responding to words than nonwords. When subjects were presented with stimuli derived from bad news items, however, subjects responded to words and non-words with the same speed.

We found no relationship between accuracy and the stimuli's relationship to good news or bad news. Semantic priming literature shows that this lexical decision task, in the absence of good news and bad news variables, consistently yields a main effect of faster response times to words than nonwords (Bentin & Wood, 1985). This is the same effect that we found when subjects were presented with stimuli derived from good news. The fact that this effect was not found when subjects were presented with information derived from bad news implies, as we predicted, that these words are being subjected to general inhibition. Our results deviated from our predictions in that non-word stimuli derived from unfavorable information were not subjected to inhibition. This implies that when the subject encounters these stimuli, the connection between the non-word and the unfavorable information from which it was derived is not recognized.

Discussion

General Discussion. In our first study, we found that participants continued to update their beliefs asymmetrically under low cognitive load and no cognitive load. When participants were under high cognitive load at the moment that we asked them to report their posteriors, however, participants updated their beliefs symmetrically. Paradoxically, putting participants in an environment in which they have less mental processing power available to devote to the task resulted in more precise processing of the information. This pattern of results shows that asymmetric belief updating does not occur through some passive, instinctive process. Our results indicate that asymmetric updating is mediated by a process that acts at the moment of information retrieval and that this process is an active one that can be disrupted by limiting the mental processing resources available to the subjects.

Our second experiment was intended to shed light on the role, if any, that inhibition plays in asymmetric belief updating. We hypothesized that the failure to use unfavorable information when forming beliefs could be the result of general inhibition of all unwanted information. As such, we predicted that if we asked subjects to assess the word/non-word status of stimuli derived from both favorable and unfavorable information that they had received, the asymmetric belief updating mechanism would inhibit their ability to recall words tied to unfavorable information. This pattern of performance would indicate that the mechanism mediating asymmetric belief updating generally inhibits any thought relating to the undesirable event information. Our results confirmed our hypothesis. Subjects were faster at recognizing words than non-words when stimuli were derived from good news items, but not when the stimuli were derived from bad news items. These results imply that inhibition of unfavorable information is not specific to the integration of the base-rate information with belief formation, but rather any processing of unfavorable information is subject to inhibition.

Now, recall the study presented earlier in which TMS was used to examine and disrupt belief formation. Here, they found that TMS, when applied to the left IFG, ameliorated asymmetric belief updating. The left IFG's previous implications in inhibitory processes suggested that inhibition plays a role in this phenomenon. In addition, the fact that the effect disappeared when the activity of this region was disrupted implied that the process mediating asymmetric belief updating was likely an active one. Our goal was to build on this previous work to explore when and how asymmetric belief updating occurs as well as its relationship to inhibition. The cognitive load study yielded results that affirmed our hypotheses that asymmetric belief updating is an active process at the moment of information retrieval. The results of the lexical decision task study suggested that all unfavorable information is generally inhibited, thus supporting our hypothesis and refining our understanding of the role that inhibition plays in belief updating.

Future Directions. These two experiments probed different aspects of asymmetric belief updating. The cognitive load study was intended to indicate whether asymmetric belief updating is an active process and whether the mechanism that mediates it is acting at the moment of information retrieval. The lexical decision task experiment explored whether or not inhibition from asymmetric belief updating would ameliorate priming effects, giving us insight into whether this phenomenon represented general or specific inhibition. The results of these experiments give valuable insight into how the mechanisms of asymmetric belief updating function. From our results, it appears that asymmetric belief updating does occur at the moment of information retrieval, that it is the result of an active cognitive process, and that the inhibition associated with this phenomenon extends to the use of unfavorable information in semantic retrieval tasks. While these are informative and exciting results, by no means have we solved the mysteries of asymmetric belief updating. There are a number of promising future directions to explore.

One potentially fruitful area of study could be that of those who, whether intrinsically or as a result of mental illness, are pessimistic in nature. Recent research shows that individuals who suffer from anxiety and depression do not to show asymmetric belief updating (Korn et al, 2014). In addition, this research shows that individuals with anxiety and depression update their beliefs with a pessimism that grows as a function of the severity of their symptoms of mental illness. Studying asymmetric updating in mentally ill individuals potentially offers a major opportunity to gain insights into both the processes that underlie asymmetric belief updating and the factors that play a role in these mental illnesses. Particularly, it would be fascinating to examine the brain activity of these individuals in comparison with that of healthy individuals. We would guess that those who suffer from mental illness would show decreased activity in the left IFG when reporting base-rates. If this were the case, this would affirm some of the results we have seen in this paper. Regardless of the specific results of these studies, gaining a clearer understanding of the mechanisms that give rise to these biases in belief updating could prove immensely valuable in developing better informed treatment modalities for those suffering from mental illness.

Another potential avenue for further research would be examining the effects of cognitive load specifically at the moment of encoding base-rate information. That is, putting subjects under cognitive load while receiving rather than retrieving base-rate information. While our work indicated that the mechanism of asymmetric belief updating is acting at the moment of retrieval, there very well may be a more complicated story that involves mechanisms at both moments. Whether the results of an experiment of this kind support our hypothesis or refute it, they would provide valuable insight.

Exploring this mechanism of asymmetric belief updating is difficult. Each experiment sheds a narrow sliver of light on a process that is likely much more complicated than we know. The two studies that we ran provided new insights into how and when asymmetric belief updating occurs. There are many more angles of this mechanism to examine and it is exciting to consider what each future experiment might uncover.

Acknowledgments

I would like to express my deepest gratitude do Matthew Jordan, Professor Laurie Santos, and the Yale Comparative Cognition Laboratory; their brilliant ideas, mentorship, and guidance were essential to the success of this project. I am thankful to the Silliman College Mellon Forum for creating a space in which the work of Silliman seniors could be shared and discussed. Finally, I am immensely grateful for my friends and family who provided me with unconditional love and support throughout my time at Yale.

References

Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(3), 336.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and clinical Neurophysiology*, *60*(4), 343-355.

Cools R, Clark L, Owen AM, Robbins TW (2002) Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. J Neurosci 22:4563–4567.

Cross, K. P. (1977). Not can, but will college teaching be improved?. *New Directions for Higher Education*, *1977*(17), 1-15.

Eli D, Rao J-M (2012) The Good News-Bad News Effect: Asymmetric processing of objective information about yourself. American Economic Journal: Microeconomics 3:114138

Hoorens, V. (1995). Self-favoring biases, self-presentation, and the self-other asymmetry in social comparison. *Journal of Personality*, *63*(4), 793-817.

Hoorens, V., & Buunk, A. P. (1992). Self-serving biases in social-comparison-illusory superiority and unrealistic optimism. *Psychologica belgica*, *32*(2), 169-194.

Korn, C. W., Sharot, T., Walter, H., Heekeren, H. R., & Dolan, R. J. (2014). Depression is related to an absence of optimistically biased belief updating about future life events. *Psychological Medicine*, *44*(03), 579-592.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, *108*(3), 480.

McCormick, I. A., Walkey, F. H., & Green, D. E. (1986). Comparative perceptions of driver ability—a confirmation and expansion. *Accident Analysis & Prevention*, *18*(3), 205-208.Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of experimental psychology*, *90*(2), 227.

Mobius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2011). *Managing self-confidence: Theory and experimental evidence* (No. w17014). National Bureau of Economic Research.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist*, *38*(1), 63-71.

Sharot T, Korn C-W, Dolan R-J (2011) How unrealistic optimism is maintained in the face of reality. Nat Neurosci 14:1475–1479.

Sharot, T. (2011). The optimism bias. *Current biology*, *21*(23), R941-R945.

Sharot, T., Kanai, R., Marston, D., Korn, C. W., Rees, G., & Dolan, R. J. (2012). Selectively altering belief formation in the human brain. *Proceedings of the National Academy of Sciences*, *109*(42), 17058-17062.

Sharot, T., Riccardi, A. M., Raio, C. M., & Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. *Nature*, *450*(7166), 102-105.

Stajkovic, A. D., Locke, E. A., & Blair, E. S. (2006). A first examination of the relationships between primed subconscious goals, assigned conscious goals, and task performance. *Journal of Applied Psychology*, *91*(5), 1172.

Suls, J., Lemos, K., & Stewart, H. L. (2002). Self-esteem, construal, and comparisons with the self, friends, and peers. *Journal of personality and social psychology*, *82*(2), 252.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12*(2), 257-285.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in cognitive sciences*, *10*(4), 141-142.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2008). A cognitive load approach to lie detection. *Journal of Investigative Psychology and Offender Profiling*, *5*(1-2), 39-43.

Wojnowicz, M. T., Ferguson, M. J., Dale, R., & Spivey, M. J. (2009). The self-organization of explicit attitudes. *Psychological Science*, *20*(11), 1428-1435.

Zuckerman, E. W., & Jost, J. T. (2001). What makes you think you're so popular? Self-evaluation maintenance and the subjective side of the" friendship paradox". *Social Psychology Quarterly*, 207-223.

Appendix 1
Events and their statistical base-rates:

STUCK IN TRAFFIC	**69**
CUT BY KNIFE	**65**
CARD FRAUD	**62**
LOSING HOUSE KEYS	**61**
SPORTS ACCIDENT	**60**
MOUSE IN HOUSE	**58**
SKIN BURN	**58**
KNEE ARTHRITIS	**54**
$30,000 IN DEBT	**56**
CHEATED ON BY SPOUSE	**52**
STUNG BY BEE	**45**
HERNIA	**43**
HOUSE OR CAR VANDALIZED	**43**
FLEAS OR LICE	**42**
TRAFFIC OR HOUSE ACCIDENT	**42**
BONE FRACTURE	**41**
DEATH BEFORE 80	**41**
INSOMNIA	**40**
OBESITY	**36**
HOUSE BURGLARY	**32**
CANCER	**30**
BACK PAIN	**70**
THREE WEEKS IN HOSPITAL	**58**
VACATION CANCELLATION	**58**
BICYCLE THEFT	**54**
PASSENGER IN CAR ACCIDENT	**53**
DIVORCE	**50**
BULLYING AT WORK	**46**
MISS A FLIGHT	**45**
ARTERIES HARDENING	**43**
THEFT FROM PERSON	**42**
OSTEOPOROSIS	**41**
WITNESS TRAUMATIZING ACCIDENT	**40**
SEXUAL DYSFUNCTION	**37**
HEPATITIS	**36**
SEVERE TEETH PROBLEMS	**35**
BEING FIRED	**62**
LOSING WALLET	**60**
HOUSEHOLD ACCIDENT	**59**