Comparing the Believability of Emotion and Event Utterances

Julie Nam (julienam@stanford.edu)
Department of Symbolic Systems, Stanford University
Stanford, CA 94305 USA

Abstract

People often make utterances about the emotional state of someone or an event that happened in order to explain that person's behavior. While there exist theories on behavior explanations, there is to my knowledge no study that focuses on the difference between emotion and event utterances that serve as behavior explanations. However, emotion as a mental state has a unique characteristic of being accessible only by the subject. In this paper, I hypothesize that there must be a difference between the degree of believability of emotion utterances and that of event utterances. I describe a model of behavior explanation that incorporates the idea that utterances may be accepted as credible or discarded. I conduct experiments to show that my model's predictions fit behavioral data most closely when its noise parameters reflect my hypothesis. I discuss the results, implications, and possible next steps.

Keywords: philosophy of mind; behavior explanation; mental states; utterance credibility

Introduction

A scene in Legally Blonde (2001) features Warner breaking up with Elle. As Elle cries loudly in the restaurant, Warner says to people around them, "Bad salad." People continue to stare in disbelief. Here, even though Warner does not say, "This woman is crying because she just had a bad salad, not because she is saddened by the breakup," we can infer the intent from context. Also, it is clear that few people will believe salad is the actual cause of her behavior. The scene shows that people offer explanations for other people's behavior, and that they are not always taken to be true. In this paper, I offer an introductory model and experiment that explore the degree with which people accept causal information about others' behaviors. The study builds on several concepts.

The foundational assumption is that behaviors are not self-generating and can be explained by other elements. This is why it is possible to infer certain states about the person or the world based on her action (e.g. she keeps staring at her watch, so she must be in a rush). While there have been a lot of work done on causal explanations and inference (Lombrozo 2012), Malle has specifically focused on behavior explanations. He posits that unintentional behavior, such as displays of emotion, can be explained by causes (Malle 2007).

My study assumes that there are largely two ways people can come up with the cause other than hearing it directly from the subject. One is through the information that others give, and another is through inference. Even when others offer explanations, people may choose to make the inference themselves if they mistrust the information given. Others

have studied the way listeners revise world knowledge based on others' utterances (Degen 2015). The general consensus is that people don't always take others' utterances as given. In this study, therefore, I assume that people perceive noise from utterances that are explanations of behavior, the level of which represents how much people self-infer the cause of the behavior as opposed to believing the utterance they hear.

The most important part of my study is on the difference between emotion and event utterances. Often, a state of the world causes in us a certain emotion, which causes us to show a certain display. When we observe somebody's display, therefore, we are led to think that there is some emotion and event behind it. In the study, I examine whether people's level of mistrust in others' utterances about the subject's emotion is different from when the utterances are about an event. Specifically, I hypothesize that people assume more noise in emotion states than in events.

The basis for this hypothesis ranges from intuitive to theoretical. Intuitively, people are less likely to make an utterance about an event that is false, because whether something happened or not is relatively easy to verify or observe. However, a claim about how the subject is feeling is more prone to be a false inference. Descartes, a Cartesian dualist, claimed that the mind is a private realm accessible only to the subject (Descartes 1641). Signs that people perceive mental states like emotion to be only self-accessible can also be found in certain languages. In Japanese, only one's own psychological states can be reported directly; it is not acceptable to say, for example, "she is sad" (Majid 2012). If people think that there is no way for a third party to *know* how someone is feeling, it seems natural that they trust an information less if it is about someone else's feelings than if it is about something that happened. However, this perspective is not unchallenged; philosophers such as B.F. Skinner have endorsed the behaviorist view that mental states are also observable (Skinner 1976). In this way, the results and development of this study have further implications in the philosophy of mind.

Model

The primary model is based on a simplified world where the only emotions are happy and sad, the only events are having won a prize and one's dog having died, and the only behavior displays are dancing and shedding tears. Furthermore, I only examined situations where either emotion or event is "compatible" with the display. The assumption here is that "happy" and "prize" are compatible

to "dance," while "sad" and "died" are compatible to "tears." Experiment 1 confirms the validity of this categorization. This narrows down the possible {emotion, event, display} combinations to the following: {sad, prize, dance}, {happy, died, tears}, {happy, died, dance}, {sad, prize, tears}.

In each trial, my model on emotion takes as input three pieces of information: utterance about the subject's emotion state, the observed behavior of the subject, and noise level for emotion utterances. The noise level reflects the degree to which people reject an utterance about someone else's emotion state. In forming a belief about the subject's current emotion, the model simply accepts the utterance with probability (1-noise). With probability (noise), the model generates an unbiased inference based on the observed behavior. For example, if the subject is dancing, the inference would be "happy" with p(happy|dance) and "sad" with p(sad|dance). Finally, the model predicts the probability with which the final belief matches the utterance.

My model on event is parallel to that on emotion. It takes in the following: utterance about an event that happened to the subject, the observed behavior of the subject, and noise level for event utterances. The noise level here reflects the degree to which people reject an utterance about an event in someone's life. With probability (noise), the model generates an inference about which event happened to the subject given the observed behavior. With probability (1-noise), the model simply accepts the utterance. The model ultimately predicts the probability with which the final belief matches the utterance.

My hypothesis is that the emotion model would need a high value for noise in order to accurately predict behavioral data, while the event model would need a relatively low value for noise to do so. Whether this holds is discussed later in the paper.

Experiments

I first conducted Experiment 1 to elicit prior belief on the likelihood of each emotion given display as well as of each event given display. I then conducted Experiment 2 to measure people's reception of emotion and event utterances.

Experiment 1: Prior Beliefs

Methods 20 participants (8 female) were recruited via Amazon Mechanical Turk and paid $0.06 compensation. Each was asked a total of 12 questions, although only 8 were used for analysis in this study (see Table 1). For each question, the participant clicked on a point in a sliderbar that shows that numerical percentage represented by the point.

Table 1: Prior Experiment Questions

Chances someone is	given that he or she
happy	won a prize?
happy,	(his or her) dog died?
sad,	won a prize?
sad,	(his or her) dog died?
dancing,	is happy?
dancing,	is sad?
shedding tears,	is happy?
shedding tears,	is sad?

Results The mean of each conditional probability was calculated. Then probabilities of display given event were also calculated. See Figure 1 for results (the last two bars in each section are irrelevant to this particular study). Since there are only two emotions and two events, using Bayes' Theorem, the probability of display given event or emotion is equal to the probability of event or emotion given display. From the results of this experiment, therefore, I could figure out the subjects' prior beliefs as shown in Table 2. The results supported the intuitive classification of compatibility of emotions and events to displays, as the differences between the probabilities of each emotion or event given the same display were all statistically significant.
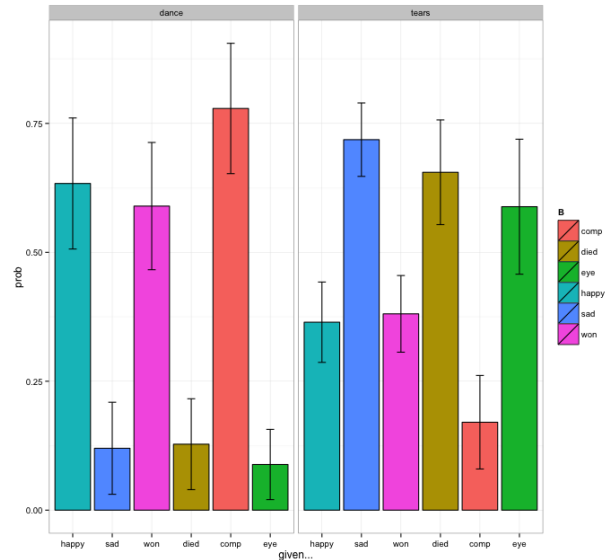

Figure 1: Prior Experiment Results in Bar Graph.

Table 2: Normalized Prior Beliefs.

p(happy | dance)	0.8407
p(sad | dance)	0.1593
p(happy | tears)	0.3366
p(sad | tears)	0.6634
p(prize | dance)	0.8218
p(died | dance)	0.1782
p(prize | tears)	0.3675
p(died | tears)	0.6325

Experiment 2: Utterance and Response

Methods 40 participants (17 female) were recruited via Amazon Mechanical Turk and paid $0.3 compensation. 20 participants were assigned two scenarios, {happy, died, tears} and {sad, prize, tears}, and 20 participants were assigned the other two scenarios: {happy, died, dance} and {sad, prize, dance}. For each scenario, participants read a 4-sentence description of the subject's behavior and utterances about it like *You run into Aaron, Bob, and Corey on the street. Aaron is dancing. Bob says to you, "Aaron's dog just died." Corey says to you, "Aaron is happy."* They were then asked to provide their beliefs about the subject's emotion state and an event that happened to the subject, e.g. *Bob and Corey leave. Doug comes by and asks, "How is Aaron feeling? What happened?" What would you say in response to the two questions?*, with drop-down menus that provide as options "happy," "sad," "other" for the question on emotion and "something good happened, e.g. he won a prize," "something bad happened, e.g. his dog died," "other" for the question on event. The order of emotion and event questions was randomized, and clicking on "other" allowed a free response.

Results In analyzing the data, I first categorized {happy, died, tears} and {sad, prize, dance} as Event Compatible prompts and {sad, prize, tears} and {happy, died, dance} as Emotion Compatible prompts. For each category of prompts, I calculated the rate in which the participant's response matched the information given by the utterance. Figure 2 shows the average rates for each category.
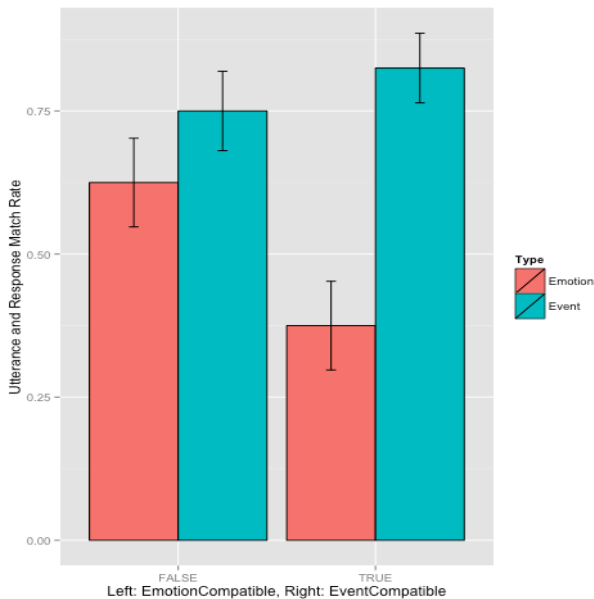

Figure 2: Average match ratings for both emotion and event given Emotion Compatible and Event Compatible prompts. Error bars are standard error over the 40 items.

Among Emotion Compatible prompts, there is no statistical difference between the match rate of emotion and that of event ($p = 0.23$). Among Event Compatible prompts, however, the match rate of emotion is significantly lower than that of event ($p < 0.0002$). Utterances about the subject's emotion were believed less than utterances about an event that happened to the subject when the uttered emotion was not very likely and the uttered event was relatively likely given the display. However, utterances about event were not believed less than utterances about emotion even when the uttered event was unlikely and uttered emotion was relatively likely given the display. This supports my hypothesis that people believe event utterances more strongly than they do emotion utterances; even when the uttered event seems incompatible with the display, people believe it to the same extent that they do a seemingly compatible emotion.

Another way I analyzed data was by comparing the match rates of emotion and event utterances when they both are compatible or incompatible with the display. Figure 3 shows this.
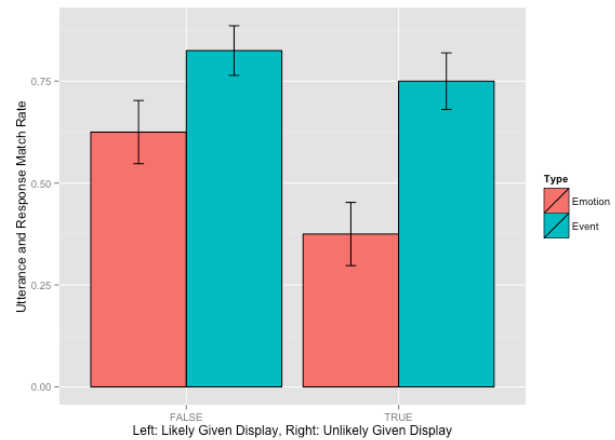

Figure 3: Average match ratings for both emotion and event given incompatible and compatible displays.

The match rate of event when the event is compatible with display is significantly higher than the match rate of emotion when the emotion is compatible with display ($p = 0.049$). The match rate of event when the event is incompatible with display is also significantly higher than the match rate of emotion when the emotion is incompatible with display ($p = 0.001$). These results support my hypothesis that people assume higher noise in emotion utterances than in event utterances in general; utterances about plausible emotions are believed less than utterances about similarly plausible events, and utterances about relatively implausible emotions are believed less than utterances about similarly implausible events.

Model Prediction

Based on the data from the two experiments, I was able to compare the model's prediction and behavioral data for each prompt. See Table 3 for details of each prompt.

Table 3: Prompt-specific Content.

Scenario #	Emotion, Event, Display
1	Sad, Prize, Dance
2	Happy, Died, Tears
3	Happy, Died, Dance
4	Sad, Prize, Tears

See Figure 4 for the emotion model's predictions of match rates for each scenario with both high noise (0.8) and low noise (0.2). See Figure 5 for a representation of results from actual data. They show that the emotion model more accurately fits behavioral data with high noise than with low noise.
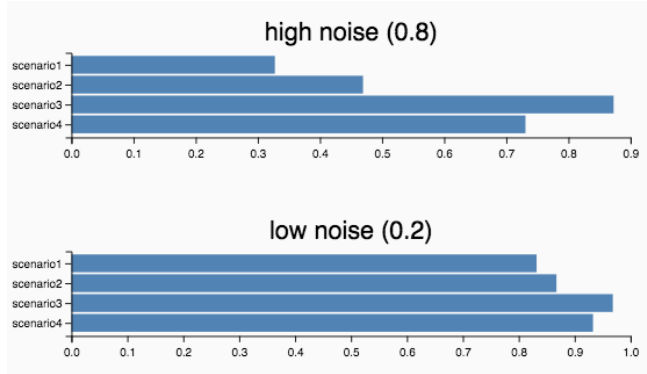

Figure 4: Emotion model's prediction of match rate for each scenario according to noise level.
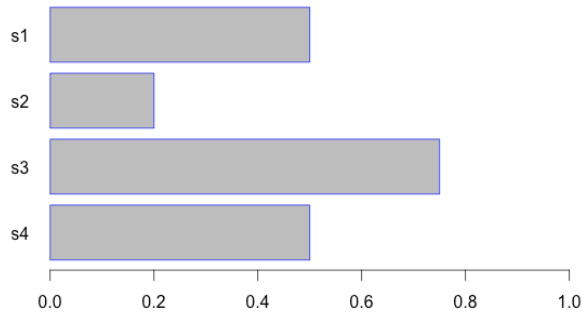

Figure 5: Emotion match rate for each scenario from experiment data.

See Figure 6 for the event model's predictions of match rates for each scenario with both high noise (0.8) and low noise (0.2). See Figure 7 for a representation of results from actual data. They show that the event model more accurately fits behavioral data with low noise than with high noise.

Overall, my models correctly predict that emotion utterances have higher noise than event utterances, as its results are closest to actual data with such setting (i.e. when noise for emotion (0.8) is bigger than noise for event (0.2)).
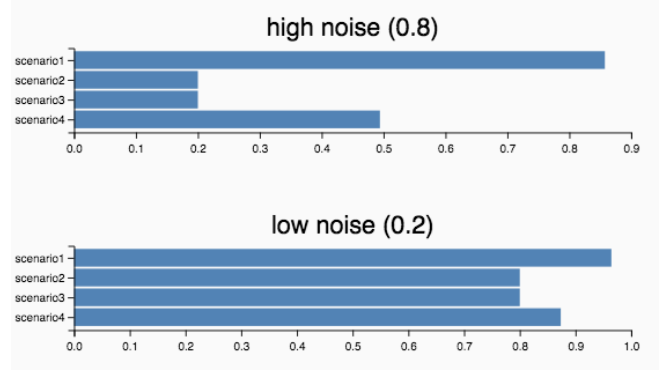

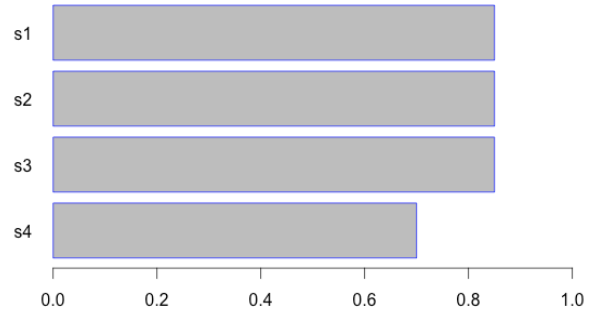Figure 6: Event model's prediction of match rate for each scenario according to noise level.

Figure 7: Event match rate for each scenario from experiment data.

Discussion

I have presented a model that predicts the rate with which people's final belief about the subject's emotion or event matches the utterance they heard about it. According to the model, there is a noise level for each type of utterance that leads people to infer the subject's state instead of simply accepting the utterance they heard. Computationally, a higher noise level leads to a lower match rate and vice versa. My model predicts behavioral data best when it assumes a higher noise level for emotion utterances than for event utterances. This means that if my model is correct, then my hypothesis that people believe emotion utterances less than they do event utterances holds. The data from my experiments, independent from the model, also supported the hypothesis as it showed significant differences in the match rates of emotion and event utterances.

This has important implications for the philosophy of mind. When it comes to the debate between dualists and behaviorists on whether mental states are accessible, my model supports the dualist's view that emotions belong to a private realm. While behaviorists would argue that mental states are not any less observable than events, my results show that there are distinct differences between emotions and events; since emotions are less verifiable than events, people tend to believe utterances about others' emotions less than they do utterances about events.

Further steps would include more complex versions of the model. Rather than keeping the idea that noise level is consistent across all situations, a model could hypothesize that noise level depends on how likely the information presented by the utterance is. Then, the model would introduce more parameters specific to the utterance type that influences the noise level. Having a higher number of possible events, emotions, and displays would also be an appropriate next step to make the model more realistic. Finally, a working model that produces the noise level given data, instead of having the noise value as input, could facilitate further studies in this area.

References

Degen, M. H. Tessler, & N. D. Goodman. Wonky worlds: Listeners revise world knowledge when utterances are odd. J. *Proceedings of the Thirty-Seventh Annual Conference of the Cognitive Science Society*.

Descartes, R. (1641). Meditations on First Philosophy in The Philosophical Writings of Ren Descartes, trans. J. Cottingham, R. Stoothoff, D. Murdoch, Cambridge: Cambridge University Press, vol.2, 1984.

Lombrozo, T. (2012). Explanation and Abductive Inference. *Oxford Handbook of Thinking and Reasoning*, 260-276.

Majid, A. (2012). Current emotion research in the language sciences. *Emotion Review, 4,* 432-443.

Malle, B. F. (2007c). Attributions as behavior explanations: Toward a new theory. In D. Chadee and J. Hunter (Eds.), Current themes and perspectives in social psychology, 3-26. St. Augustine, Trinidad: SOCS, The University of the West Indies.

Skinner, B. (1974). *About Behaviorism.* New York: Random House, Inc.