Running Head: LINGUISTIC REPRESENTATIONS OF PROBABILITY

How Might We Use Words like Might? Categorical Distinctions in Epistemic Modal Verbs as Seen Through Modal Judgment Tasks

Serena Meghani

Psychology Department, University of California Berkeley

Thesis advisor: Shaun O'Grady

Faculty mentor: Mahesh Srinivasan

Author note

Serena Meghani, Cognitive Science, UC Berkeley. Special thanks to Shaun O'Grady for acting as my mentor through this senior thesis, as well as Ariel Starr, Ruthe Foushee, and Professor Mahesh Srinivasan.

Abstract

Epistemic modal verbs are a special class of verbs used to express the possibility of events based on limited evidence and the speaker's knowledge of the situation. Little research has explored the relation between the use and comprehension of epistemic modal verbs and the perceived probability of events. In three online survey studies utilizing different tasks, native English speakers made judgements assessing various epistemic modal statements in describing predictions of the outcome of a random draw from varying distributions. Our results show converging evidence that generally participants accept strong modal phrases more when paired with high probability events as compared to low probability events, suggesting a categorical distinction between strong and weak modal verbs. These findings support the view that epistemic modal verbs form a natural order, and provide insights into the logical and pragmatic interpretations of modal verbs when used epistemically. Future research will look at how children develop their understanding of modal expressions and probability.

Keywords: Epistemic modality, probabilistic reasoning, Linguistic representations of probability

Linguistic Representations of Probability.

Every day, we come across many situations in which we need to express uncertain and probable outcomes. Words like “might”, “could”, “likely”, and “possible” are some of the many ways in which we express uncertainty. They are how we decide, when someone tells us “it *might* rain today”, whether we think we should grab an umbrella, or a light jacket. Expressions of uncertainty also vary person to person. Does “might” and “should” mean the same thing to one person as it does to someone else? How do we all reach consensus on the probability of events occurring based on language used?

Modern humans communicate risk, uncertainty, and probability, using linguisitic representations several times a day yet we know very little about the quantitative properties of these representations. The purpose of the current research is to investigate the relationship between linguistic expressions and simple probability judgments. Through three different surveys utilizing separate tasks, we assessed adults’ judgement of statements containing modal verbs in reference to the possibility of an outcome randomly drawn from binary distributions of green and purple marbles.

Humans need to understand expressions of uncertainty in order to make decisions, especially in situations with probabilistic outcomes. Although research on decision making has highlighted flaws in human reasoning when reasoning about conditional probability (Kahneman & Tversky, 1973) very little research has explored judgments about simple probabilities based on binary proportions. These types of tasks are often presented as predictions of the outcome of a random draw. Piaget & Inhelder (1975) first used the random draw task in their developmental

study of children's understanding of chance. Adults perform at ceiling on random draw tasks designed for children (Chapman, 1975) and their performance on a 2-alternative forced-choice probability discrimination task suggests that adults are capable of accurately discriminating binary proportions (O'Grady, Starr, Griffiths & Xu, submitted). The purpose of the current research is to investigate the relationship between linguistic expressions and binary probability judgments.

There are many communication factors and psychological implications that come with a better understanding of modality. Holmes (1982) noted that second-language learners have difficulty expressing epistemic modality due to variations in the degree of certainty expressed by each expression. Papafragou (2006) showed that epistemic modality contributes further information of the belief of truth the speaker has on the proposition. In terms of psychological implications, the development of epistemic modality has been thought to be linked to children's development of theory of mind, (i.e. the ability to understand mental representations and reason inferentially about them when thinking about and communicating to others) (Papafragou, 1998). This theory has been further developed by showing that autistic individuals, who are argued to have deficiency in theory of mind capabilities, have difficulty with epistemic modality (Papafragou, 2002). By investigating the semantic and pragmatic meaning underlying epistemic modality, we hope to gain insight into the psychological representations of uncertainty as well as the factors which facilitate or inhibit the accurate communication probability.

Modality, often defined as a speaker's subjective attitude towards or belief about a propositional statement can be communicated in many ways, including adverbs, adjectives, and

other linguistic phrases (Lyons, 1977). Most studies regarding linguistic expressions of probability focus on adverbs and phrases, as opposed to modal verbs. Of particular interest to the current discussion, people often use modal verbs to express uncertainty. The literature on the topic of modal expressions generally divides these terms into two types of modal meanings: deontic modal meanings, which refer to expressions of obligation, ability or permission, and epistemic meanings, which refer to expressions involving the possibility of events based on limited evidence and the speaker's knowledge of the situation. The underlying mental processes related to the two meanings differ, a supposition that is supported by children's later acquisition of epistemic meanings as opposed to deontic meanings (Papafragou, 1998). A third class of modal meanings, often associated with epistemic modal meanings, is alethic modal meanings, which refer to logical necessity, probability and possibility. Epistemic uncertainty is thought to be associated with confidence statements ("I am reasonably certain"), while aleatory uncertainty is thought to be associated with likelihood statements ("I believe it is fairly likely") (Ülkümen, Fox, & Malle, 2016). The main difference between epistemic and alethic modality is that alethic modal meanings reflect real-world truths, while epistemic modal meanings reflect what one believes to be true (Frawley, Eschenroeder, Mills, & Nguyen, 2006). This study will focus on the acceptance and use of epistemic and alethic modal meanings.

Modal verbs are polysemic, meaning that the same set of words are used in both epistemic and deontic meanings. The meaning of modal verbs vary in different contexts. In a study analyzing a large corpus of Technical English (i.e. science research papers) and General English (i.e. literary works), Jaime & Guillot (2015) found that in technical writing, modal verbs were

used to express possibility most of the time (83%) and in general writing, modal verbs were used to express possibility just as much as ability (~32%). Furthermore, in general contexts, there was a greater variety of meanings used for the same word, with would, may, might, can, and could all having at least two different semantic meanings, one of which was probability/possibility (Jaime & Guillot, 2015). This suggests that the two meanings are both used with similar frequency in everyday speech, and therefore may be hard to disentangle. Some accounts suggest that the later development of epistemic modal meanings in children as compared to deontic meanings is due to the polysemic nature of these expressions (Sweetser, 1990). Coates (1987) found that children starting from 8 years of age show deontic and epistemic differences in sorting modal expressions, but that the meaning of these modal expressions becomes more precise with age until adulthood. The differences in acquisition implies that deontic and epistemic meanings have different underlying psychological factors. This study will focus on having subjects understand these modal expressions strictly with an epistemic interpretation.

Semantically, epistemic verbs concern the degree of possibility or necessity based on available evidence. There is much debate on how the epistemic scale is divided. Previous research and standard analysis suggest that modal verbs and adjectives can be organized on a continuum ranging in likelihood from disbelief to belief, and ordered by informational strength (Reyna, 1981; Horn, 1972). Other studies suggest that epistemic modality is best understood when divided into discrete categorical distinctions (Hahn & Englelmann, 2014; Rubin, 2010; Clark 1990).

LINGUISTIC REPRESENTATIONS OF PROBABILITY

Many studies in linguistics have tried to match uncertainty expressions to probability values or ranges (see Clark 1990 for an on overview). One common method utilized are scaling tasks, in which subjects sort and rank linguistic expressions either based on similarities of the sentences to one another or based on the magnitude of likeliness to occur, both of which were used by Reyna (1981). However, these tasks have low ecological validity and assess the modal expressions without the direct assignment of objective probabilities. The tasks can be seen as ecologically invalid because they vary in terms of context (Clark, 1990). That is, verbal uncertainty terms are contextually dependent, and, as such, a participant reading a phrase such as "it is possible that it will happen" is heavily dependent on what "it" is (Reyna, 1981). Similarly, in a study conducted by Mauboussin and Mauboussin (2018) participants were asked to match various probabilistic expressions to percentages. They found less variance for strong modal expressions on both positive and negative ends (i.e. "always"/"certainly" and "rarely"/"never") and greater variance for modal expressions related to "possible" and its variations. However, the survey was worded as "If a future event is X to happen, what percentage of the time would you estimate it ends up happening?" with X being different modal expressions. Again, this could be dependent on what the participants associate with the "future event" based on context. This idea of the importance of context was seen in an earlier study conducted by Beyth−Marom (1982) which compared context-free and context-dependent tasks, and found higher ranges of probability assignments in 14 of the 30 expressions in context-dependent tasks. Both of these tasks are just ways of summarizing information, since there is a lack of direct assignment of probabilities or lack of direct objective contexts (Clark, 1990). To further understand the semantics of epistemic modal verbs, the current study will pair modal expressions with a

situational context of the chances of obtaining a green marble from a group of green and purple marbles. This will offer greater ecological validity, as well as a way to assess the modal expressions in relation to a direct assignment of objective probabilities.

Alongside the question of the semantics of epistemic modal verbs is the idea that the boundaries between epistemic states are unclear and vary based on individual perception. According to Clark (1990), studies looking at quantifying epistemic modality consistently find low within-subject variability and high between subject-variability. This is important because it suggests that each individual has a different underlying mental representation and understanding of modal verbs that is consistent within how he or she uses these expressions, but vary in how others use the same expression. In one study, four annotators individually analyzed the use of epistemic modal statements in 80 articles from a New York Times dataset. Using this method, Rubin (2010) found that the same expressions had different underlying conceptions of modal certainty for each participant. Together, these studies suggest that more research is required in understanding how epistemic modality is understood and communicated.

The other means of analyzing modal expressions is from a pragmatic interpretation. Pragmatically, this epistemic scale gives rise to scalar implicature. Scalar implicature is the idea that we can implicitly understand the strength of the speaker's evidence, by the informational strength. Consider the following statements:

(1) It may snow today.

(2) It has to snow today.

Note that example (2) is stronger than example (1) in the belief direction. According to Grice's Quantity Maxim (1989), communication should be as informative as possible for the current exchange. Therefore, if someone says statement (1), there is an understanding of uncertainty. This gives rise to the following statement:

(3) It does not have to snow today.

The computation of scalar implicature therefore involves understanding a weak scalar term to imply the negation of the stronger scalar term. Furthermore, the listener has a general understanding that if the speaker was more certain, they would have said statement (2) instead of statement (1).

Expressions of epistemic modality are interesting in this regard because what is logically true may not be pragmatically felicitous (i.e. true based on the context). Developmental studies show that semantic and pragmatic aspects of modal linguistic expressions develop at different stages. O'Neill and Atance (2000) investigated children's use of modal utterances in the CHILDES database and found that children typically use modal expressions for predicting the outcome of a future event by the age of 3.5. Children seem to have an understanding of relative modal strength and semantic understanding of epistemic modal verbs between 5 and 6 years (Hirst & Weil, 1982). Ozturk and Papafragou (2015) found that children between 4 and 5 have mastered some aspects of epistemic modal semantics and can make some pragmatic inferences, but struggle with contexts involving epistemic possibility. This is also seen in other studies that find it easier for children to deduce necessary conclusions than possible conclusions when dealing with a logical interpretation of epistemic modal verbs (Noveck, 2001). Together, these

studies suggest that the development of scalar implicature arises from the transition of logical interpretations to pragmatic interpretations. However, more research is needed in understanding the semantics-pragmatics interface of epistemic modality. The current study seeks to further investigate whether people view weak modal verbs pragmatically or logically.

Other studies have begun to look at the relationship between linguistic representations and probability with similar scenario-based designs. However, our study differs in two important ways: in the distribution of probabilistic outcomes and in the choice of epistemic modal adjective. Previous studies have looked at the acceptance of weak modal expression in conditions with absolute outcomes (Hirst & Weil, 1982; Noveck, Ho, & Sera, 1996; Ozturk & Papafragou, 2015). However, there is a lack of research on the acceptance of weak modal expressions in cases of probabilistic outcomes. When probabilistic outcomes are used, they are either limited or unclear probabilities, and compare only a necessary conclusion with one plausible conclusion. Noveck (2001) began to look at probabilistic conclusions by including a condition with a 50% chance of a possible conclusion, and comparing it to an absolute condition (i.e. 100% chance). Lassiter and Goodman (2014), on the other hand, compared the acceptance of different modal adjectives when comparing valid and contingent arguments (i.e. conclusions that are 100% true and conclusions that might be true). They found that for weak modal frames like “possible” and “probable” there was high acceptance for both valid and contingent conditions, but for strong modal frames like “necessary” and “certain” there was a high acceptance for valid conclusions, but a low acceptance for contingent conclusions. While this study looked at contingent probabilistic conclusions, the magnitude of likelihood was not specified, which could influence

how acceptable subjects found the different modal expressions. Similar to these studies, our study will look at absolute, necessary outcomes by looking at trials with 100% probabilities, but will also include other clearly-defined distributions.

Secondly, most of the previous literature that studied people's understanding of modal expressions focus on epistemic modal adjectives and adverbs, as opposed to modal auxiliary verbs. As mentioned before, Holmes (1982) noted that second-language learners have difficulty expressing epistemic modality due to variations in the degree of certainty expressed by each expression. There are a lot fewer modal verbs than modal adjectives and adverbs in English, and the modal verbs tend to be less specific than some of the phrases used. The limited number of alternative verbs might affect how people map them on to probabilities. Furthermore, modal verbs might have different linguistic properties as compared to other modal expressions. While epistemic modality in the form of adverbs and adjectives have probabilistic semantics, it is still not known whether epistemic modal verbs share this same property given their lexical ambiguity (Papafragou, 2006). Given the high usage of modal verbs in describing uncertain conditions, it is important to investigate how these expressions are used and understood. Thus, the current study seeks to expand the probability distributions and modal verbs in question.

Our current study investigated the relationship between predictions about the outcome of a random draw from different binary distributions and the choice of modal verbs used to describe the likelihood of the outcome. We propose that adults are capable of communicating probability linguistically through the use of modal expressions. In Experiment 1, participants were asked to rate the usefulness of modal verb phrases in describing the prediction of a random draw outcome

for various simple distributions. In Experiment 2, we sought to untangle the logical and pragmatic interpretations of modal verbs by implementing a survey similar to the first, with the exception that participants were now told that a friend will describe the chance of a random draw outcome, and the participant must rate the usefulness of the friend's statement after seeing the distribution of marbles themselves. Finally, Experiment 3 asked participants to create their own distributions that correspond to the six modal verbs we have been investigating, to better understand how people associate modal verbs with probabilities in a modal choice task. This survey also asked participants to rank the modal verbs, so that we can see the variation of ordering in these verbs in a scaling task.

Experiment 1: Methods

Participants

Fifty-three native English speakers were recruited for an utterance acceptance task using Amazon's Mechanical Turk and Qualtrics. Twelve participants were excluded (6 failed to complete the study and 6 failed attention check trials). We planned to recruit 40 participants but our final sample size is N = 41 (22 female; Mean age = 32.43; SD = 9.19). Before beginning the online experiment, participants provided informed consent by reading and accepting an online consent form approved by the UC Berkeley Committee for the Protection of Human Subjects (CPHS).

Material

Images were created using Blender 2.72, 3D animation software. Each image consisted of a gumball machine next to a group of green and purple marbles organized into 7 ratio bins. Table

1 lists the contents of each image. The positions of the marbles in each image were be randomly generated and .jpg images were rendered for each ratio listed in table 1. Figure 1 shows an example trial image.

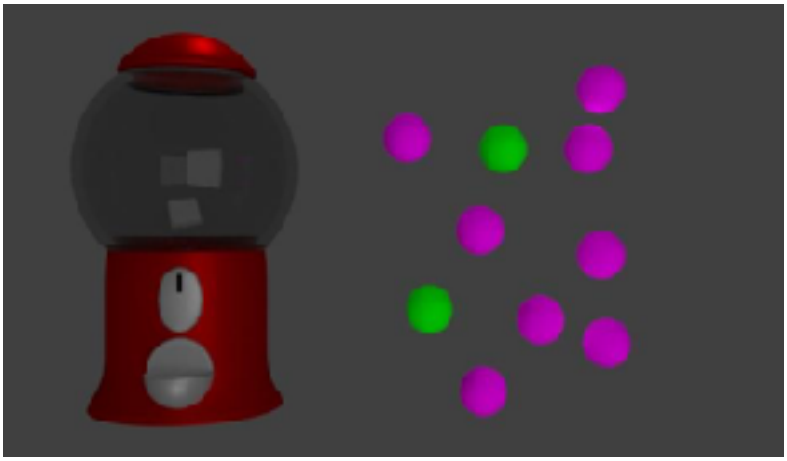

Figure 1. Example Trial Image depicting the gumball machine, and a group of 10 green and purple marbles. Proportions of the green and purple marbles varied by trial.

Modal sentences express varying degrees of epistemic meaning. For each of the images, the participant was asked to rate the usefulness of statements of the form "I X get a green marble.", where X is an epistemic modal auxiliary verb. Participants viewed each ratio image along with the following modal verbs: can, may, could, might, should, and will. These verbs were taken from The Grammar Book: an ESL/EFL Teacher's Course (Celce-Murcia, Larsen-Freeman, & Williams, 1983), a text used to teach grammar to English as a Second Language Learners.

Table 1

Proportions presented in each trial of Experiment 1.

Probability of Green	Number of green	Number of purple	Total
1.00	10.00	0.00	10.00
0.90	9.00	1.00	10.00
0.80	8.00	2.00	10.00
0.50	5.00	5.00	10.00
0.20	2.00	8.00	10.00
0.10	1.00	9.00	10.00
0.00	0.00	10.00	10.00

Note. Ratios of Proportions are rounded to 2 digits.

Procedure

Participants were instructed to rate the usefulness of various phrases in describing the chances of randomly drawing a green marble from the gumball machine images. For each trial, participants were presented with an image and a modal sentence. Participants rated the usefulness of the modal sentence in describing the image on a 7-point Likert scale, where seven is very useful and one is not useful/false. Above each image are the instructions “Imagine that you are trying to tell another person about their chances of getting a green marble. On a scale of 1-7, how useful would the following phrase be to this other person?” Each image was paired with every possible modal expression and presented in a randomized order. In total, participants were shown 7 images, and completed 42 Likert scale ratings.

Data analysis

Analyses

Analyses were conducted using the lmer package for mixed effects regression (Bates, Maechler, Bolker, & Walker, 2015) written for the R statistical programming language (R Core Team, 2008). We conducted comparisons of linear regression models with mixed effects

predicting likert scale rating from the interaction of modal sentence and the proportions of marbles in presented images using Chi square tests.

Results

The full model with interactions between modal strength and displayed probability had the best fit to the data ($R^2 = .47$, $F(11,1710) = 139.55$, $p < .001$). Inspection of the coefficients presented in table 2 revealed that the image probability had a large, positive, main effect suggesting that acceptance of modal terms tended to increase when modal expressions were presented along side images with a high probability of yielding a green marble. However, the interaction between modal expressions and probability reveals that this increase was greater for strong modal verbs such as 'will' and 'should' compared to weak modal expressions such as 'might' and 'may'. This interaction can be seen in the average acceptance ratings presented in Figure 2.

Table 2

Regression table for full model with interactions between probability and modal expression.

Predictor	*b*	95% CI	*t* (1710)	*p*
Intercept	2.47	[2.18, 2.77]	16.24	< .001

Modalcould	0.22	[− 0.21, 0.64]	1.01	.314
Modalmay	0.10	[− 0.32, 0.52]	0.46	.643
Modalmight	0.24	[− 0.18, 0.67]	1.13	.258
Modalshould	-1.43	[− 1.85, −1.01]	-6.63	< .001
Modalwill	-1.61	[− 2.03, −1.19]	-7.47	< .001
ProbabilityGreen	3.53	[3.05, 4.01]	14.52	< .001
Modalcould × ProbabilityGreen	-0.48	[− 1.15, 0.20]	-1.38	.167
Modalmay × ProbabilityGreen	-0.74	[− 1.41, −0.06]	-2.14	.032
Modalmight × ProbabilityGreen	-0.97	[− 1.64, −0.29]	-2.82	.005
Modalshould × ProbabilityGreen	1.71	[1.03, 2.38]	4.96	< .001
Modalwill × ProbabilityGreen	1.11	[0.44, 1.78]	3.23	.001

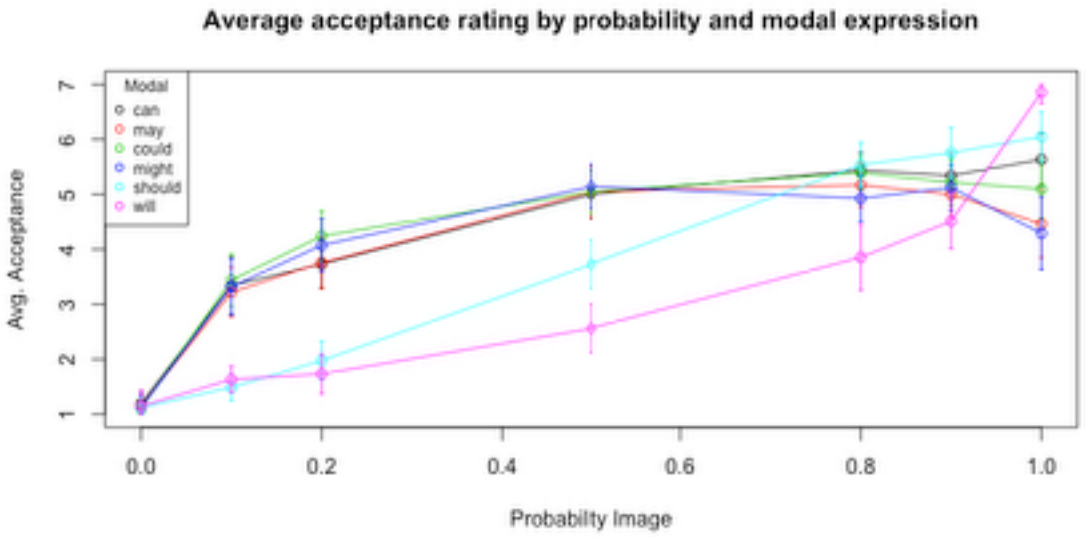

Figure 2. Mean Acceptance rating by probability and modal expression. Error bars represented bootstrapped 95% CIs.

Discussion

Results from Experiment 1 indicated that people give higher acceptance scores when strong modal expressions such as 'will' and 'should' are paired with high probability compared to low probability events. Although acceptance of weak modal when paired with low probability events was greater than acceptance of strong modals paired with the same probabilities, acceptance of weak modal expressions tended to asymptote with higher probabilities. This latter finding is quite surprising given that weak modal expressions intuitively seem inappropriate when referring to high probability and certain events. Apparently, at least a subset of the participants tended to use a 'logical' interpretation of the weak modal expressions rather than a more accurate, 'pragmatic' interpretation. In order to address this issue we altered the design of this experiment in Experiment 2 in order to increase the need for pragmatic accuracy by presenting the task as a communicative exchange between cooperating friends.

Experiment 2: Methods

Participants

Forty-seven native English speakers were recruited using the same methods employed in Experiment 1. Eight participants were excluded because they did not pass the attention check trials. We planned to recruit 40 participants but our final sample size is N = 39 ((16 female; Mean age = 32.36; SD = 8.58). Before beginning the online experiment, participants provided informed consent by reading and accepting an online consent form approved by the UC Berkeley Committee for the Protection of Human Subjects (CPHS).

Material

Images were created using the same methods described in Experiment 1 with a few notable changes. We decided not to include the trial with 0 green marbles since participants from Experiment 1 were unanimous in their rejection of all modal expression paired with this probability image. For each probability image and modal expression pair (36 total trials per participant) we created a trial-unique probability image rather than recycling the same 6 images for all of the trial pairs to eliminate potential bias of marble placement. We used the same modal expressions as Experiment 1.

Procedure

Participants were asked to imagine that they are trying to get a green marble from a gumball machine. Next they are told that a friend who sees the contents of the machine machine will tell them about their chances of getting a green marble by randomly drawing a marble from the machine. They are also informed that they will see the marbles in the bin after their friend informs them and that they will be asked to judge how useful their freind's statement is for describing the probability of drawing a green marble. For each probability image and modal expression pair, participants rated the usefulness of the freinds statement based on a likert scale ranging from 1 (not useful at all) to 7 (very useful).

Results

We used the same analytical methods described in Experiment 1. The full model with interactions between modal strength and displayed probability had the best fit to the data ($R^2 = .33$,

$F(11,1392) = 61.80, p < .001$) thus replicating our findings from Experiment 1. Inspection of the model coefficients presented in Table 3 revealed main effects of probability image and modal expression as well as an interaction between the two variables. Although probability image had a positive effect on acceptance ratings, this effect was greater for strong modals such as 'will' and 'should'. Figure 3 presents the mean acceptance ratings by probability image and modal expression and model coefficients along with 95% confidence intervals can be found in table.

Table 3

Regression table for full model with interactions between probability and modal expression.

Predictor	*b*	95% CI	*t* (1392)	*p*
Intercept	2.89	[2.48, 3.31]	13.61	< .001
Modalcould	0.38	[− 0.21, 0.97]	1.27	.204
Modalmay	0.15	[− 0.44, 0.74]	0.50	.614
Modalmight	0.46	[− 0.13, 1.05]	1.52	.130
Modalshould	-2.12	[− 2.71, −1.53]	-7.05	< .001
Modalwill	-2.59	[− 3.18, −2.00]	-8.61	< .001
ProbabilityGreen	1.84	[1.22, 2.45]	5.85	< .001
Modalcould × ProbabilityGreen	-0.71	[− 1.58, 0.16]	-1.59	.112
Modalmay × ProbabilityGreen	-0.66	[− 1.53, 0.21]	-1.49	.136
Modalmight × ProbabilityGreen	-1.07	[− 1.94, −0.20]	-2.42	.016
Modalshould × ProbabilityGreen	3.57	[2.70, 4.44]	8.03	< .001
Modalwill × ProbabilityGreen	3.62	[2.75, 4.49]	8.16	< .001

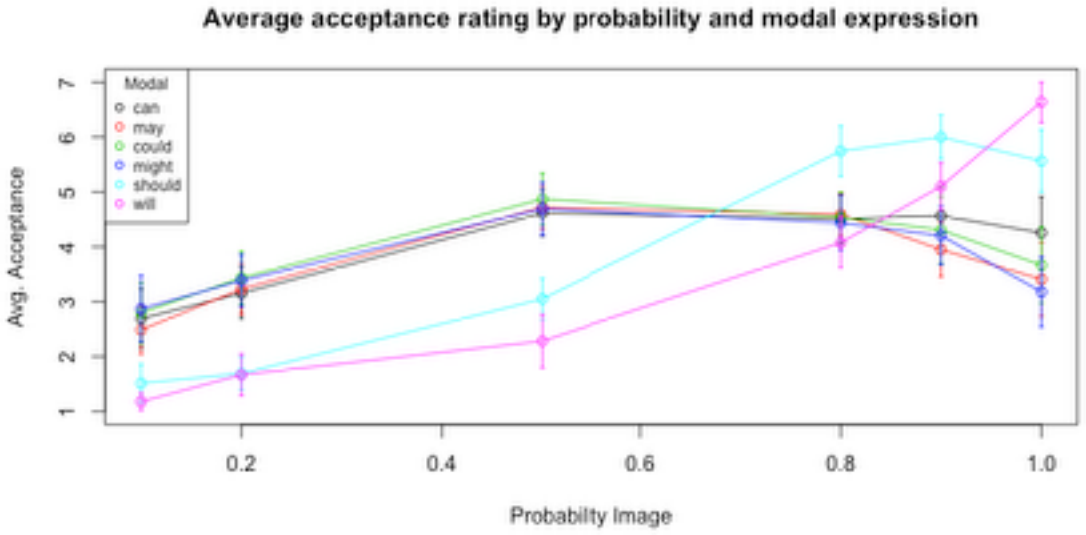

Figure 3. Mean Acceptance rating by probability and modal expression. Error bars represent bootstrapped 95% CIs.

Discussion

In Experiment 2 we replicated the major findings from Experiment 1 and refined the methods to show a clear association between strong modal expressions and higher probabilities. Although the improved methods of Experiment 2 helped to reduce the logical interpretation of weak modal expressions when paired with high probabilities, participants tended to give low usefulness ratings to weak modals more generally (Figure 3). It is possible that participants interpreted the instructions to mean ‘rate the usefulness of the phrase for informing a gamble’ rather than the ‘rate usefulness of expressing the objective probability’. Future studies should adopt a more explicit explanation in order to adress this potential ‘loss aversion’ interpretation. Another interpretation of the low acceptance of weak modals paired with low probabilities is that

weak modals may be best paired with chance probabilities while the negation of strong modals (should not, will not) may be best for communicating low probability events.

Experiment 3: Methods

While the previous 2 experiments measured how people associated probabilities with modal verbs, experiment 3 was adapted to understand how people associate modal verb phrases with probabilities. The experiment was also expanded to explicitly ask participants to rank the different modals.

Participants

Forty native English speakers were recruited using the same methods employed in Experiments 1 and 2. Fourteen participants were excluded because they did not pass the attention check trials. We planned to recruit 40 participants but our final sample size is N = 36 (12 females; Mean age = 35.15; SD = 10.27). Before beginning the online experiment, participants provided informed consent by reading and accepting an online consent form approved by the UC Berkeley Committee for the Protection of Human Subjects (CPHS).

Material

Images were created using the same methods described above, with the key difference being that the marbles were now colorless. For each modal, participants were shown the image of the gumball machine along with 10 white marbles on the side. The survey (created using Qualtrics) was made using the "Hot Spot" question type, with each marble area acting as a hot spot. Each question was repeated 3 times throughout the survey, creating 18 trials per partipant,

and 3 attention checks. On the attention check trials, participants were told the exact number of marbles they should select, to ensure that participants were paying attention to the task. The last question of the survey asked participants to explicitly rank the 6 modal expressions, and was created using the "Rank Order" question type on Qualtrics. The expressions were of the form "It X happen." where X was each of the modal verbs (will, should, might, may, could, can). These are the same modal expressions as the previous two experiments.

Procedure

Just as in the previous experiments, participants were asked to imagine that they are trying to get a green marble from a gumball machine. As in Experiment 2, participants are told that they have a friend who can see the contents of the machine and will tell them about their chances of getting a green marble by randomly drawing a marble from the machine. Participants were then asked to guess the distribution of marbles, based on the phrase their friend used. For each of the 18 trials and 3 attention checks, participants had to click between one and ten marbles. After completing those trials, participants were asked to rank the modal verbs.

Results

We used the same analytical methods described in Experiment 1. The full model with interactions between modal strength and displayed probability had the best fit to the data ($R^2 = .39$, $F(5,210) = 26.60, p < .001$) thus replicating our findings from Experiment 1. Inspection of the model coefficients presented in Table 4 revealed main effects of probability image and modal expression as well as an interaction between the two variables. Although probability image had a

positive effect on acceptance ratings, this effect was greater for strong modals such as ‘will’ and ‘should’. Figure 4 presents the mean acceptance ratings by probability image and modal expression and model coefficients along with 95% confidence intervals can be found in table. Figure 5 shows the rankings participants associated with the different modal verbs, with 1 corresponding to most-likely and 6 corresponding to least-likely. Finally, each participants’ response across the four trials is graphed in Figure 6, to analyze the within-subject variation.

Table 4

Regression table for full model with interactions between probability and modal expression.

Predictor	*b*	95% CI	*t* (210)	*p*
Intercept	5.54	[4.83, 6.25]	15.37	< .001
Modalcould	-1.35	[− 2.36, −0.35]	-2.66	.009
Modalmay	-1.42	[− 2.42, −0.41]	-2.78	.006
Modalmight	-1.31	[− 2.31, −0.30]	-2.56	.011
Modalshould	1.48	[0.47, 2.48]	2.90	.004
Modalwill	3.10	[2.09, 4.10]	6.08	< .001

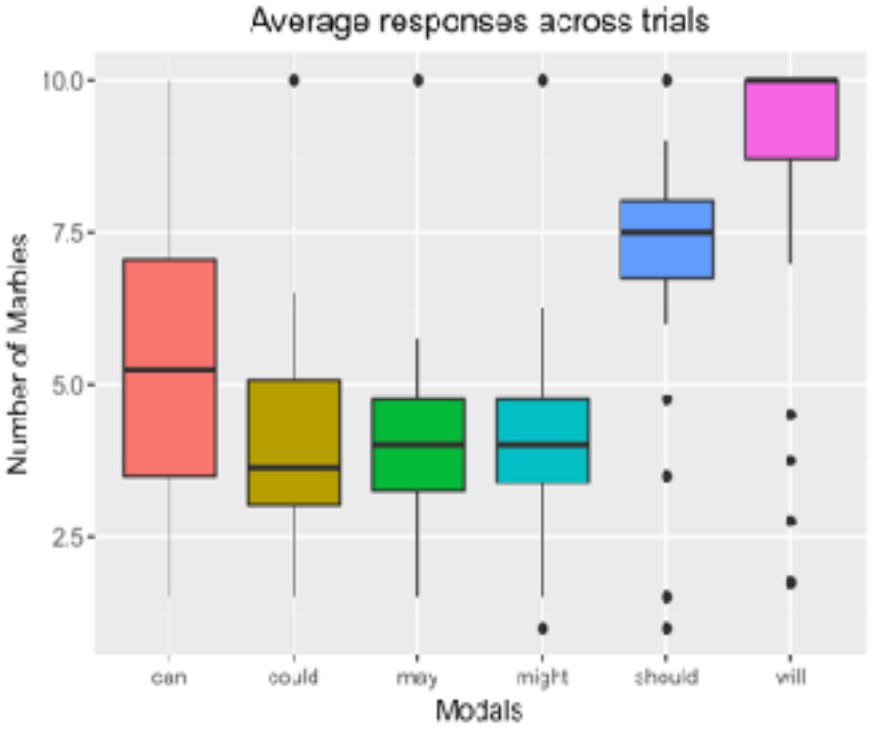

Figure 4. Mean number of green marbles marked for each modal expression across all 3 trials.

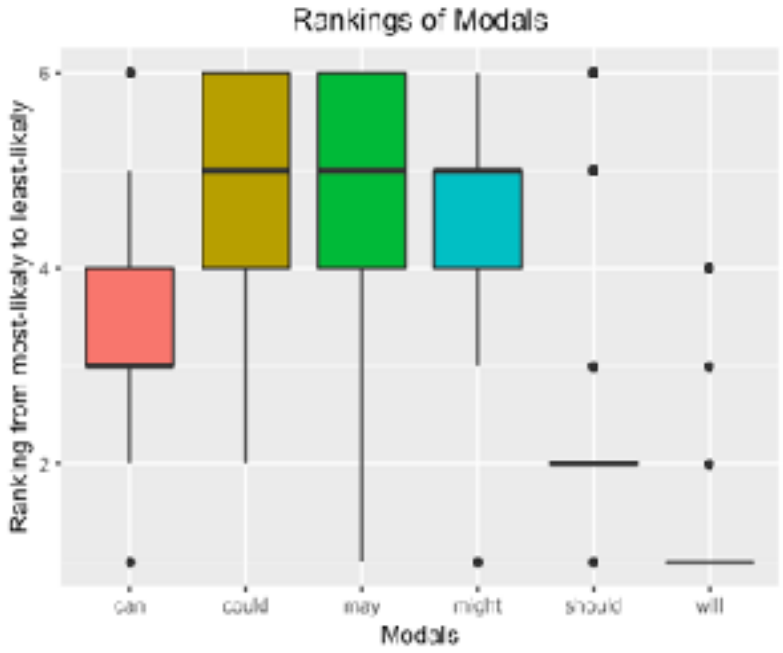

Figure 5. Mean ranking foreach of the modal expressions, with 1 being most likely, and 6 being least likely.

Discussion

As with the previous experiments, we see converging evidence of the categorical separation of modal verbs, with "will" and "should" acting as strong modals (with "will" being stronger than "should"), "might", "may", and "could" acting as weaker modals with relatively the same strength associated with the 3 verbs, and "can" being a unique modal in its own category that encompasses a range of interpretations (Figure 4 and 6). Overall, we generally see internal consistency within participants, although the exact number of marbles for the same modal verb varied trial to trial. One possible explanation for the larger variation we see in the probabilities associated with "can" is that the modal lends itself more to the multiple interpretations.

The ranking task, that models the scaling task used in linguistic studies, show the same results as our previous findings. We see the inverse of Figure 4 in Figure 5. The strong modals (i.e. "will" and "should") correspond to high acceptances in high probabilities, as well as lower rankings corresponding to beliefs of being more likely. Just as in the other tasks, the range of the strong modal expressions are smaller and less variable in the ranking task, with most of the subjects ranking "will" as the modal expression associated with the most likely to happen, and "should" following as second most likely to happen.

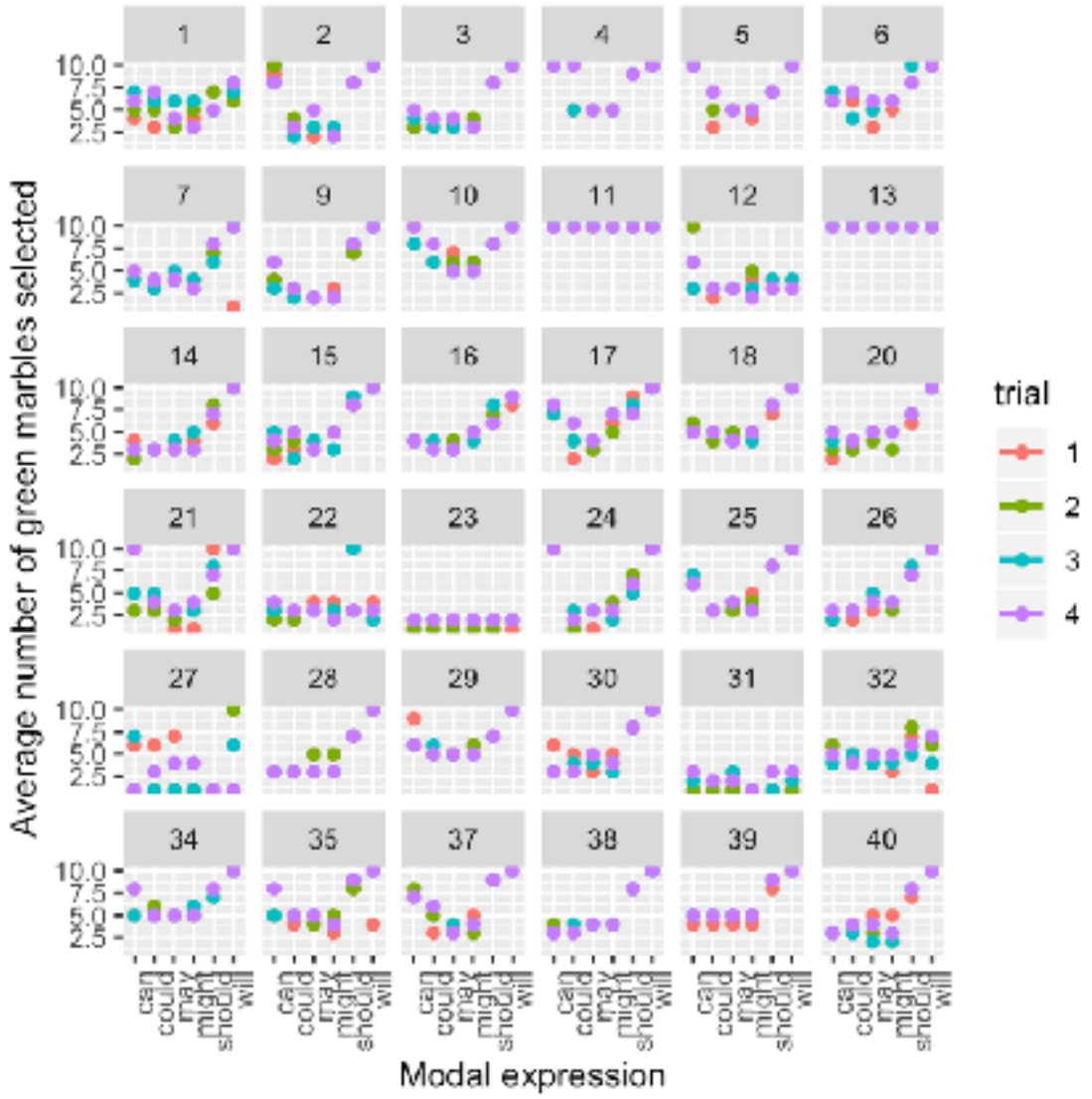

Figure 6. Results of each participant' responses for each of the 4 trials.

General Discussion

In this paper, we present converging experimental evidence of a categorical distinction between strong and weak modal verbs using modal verb acceptance and modal verb choice tasks. By having participants make judgments of various modal verbs in association with simple

probability distribution of marbles, we were able to assess the linguistic representations of probability in an ecologically valid, context-dependent task.

The results from the three survey studies suggest that there is a general consensus that people make a discrete categorical distinction between different types of modal verbs based on modal strength. Specifically, we found that there is a general distinction between strong modals ("will" and "should") and weak modals ("may", "might" and could"). Throughout all three experiments, there was a general trend that strong modals were paired with higher probabilities, regardless of the task or the direction between modals and probabilities. These results converge with the findings of Lassiter and Goodman (2014), who report high acceptances of strong modal expressions for necessary conclusions (similar to our 100% distribution). The opposite trend was seen for weak modals, although the acceptances and number of marbles associated did not drop at higher probabilities like it did in Lassiter and Goodman (2014). Similar to Mauboussin and Mauboussin (2018), there is a greater consistency of strong modal expressions and greater variation for weaker modal verbs. Results of the current series of experiments indicate that there is a distinction to be made for strong and weak modal verbs in describing high probabilistic outcomes, and to some extent, in describing low probabilistic outcomes as well.

One surprising finding from our study was the broad acceptability of the modal "can". Even in more pragmatic contexts (as opposed to logical contexts), we see a high acceptance of "can" across all probabilities. The high acceptance of "can" when paired with high probabilities suggest that the modal "can" lends itself more to multiple interpretations. For example, some may be interpreting "can" in the deontic sense, meaning they understand it as the ability to be

able to obtain a green marble. Others may be interpreting it in the epistemic/alethic sense, meaning they understand it as a way to convey the possibility of getting a green marble. Another explanation could be the difference in logical and pragmatic understanding of the word. Logically, even in cases of high probability, one "can" get a green marble. Pragmatically, however, it is assumed that "can" is of low strength as compared to other modal verbs, and should be understood as the negation of stronger modal verbs, based on scalar implicature. Future research will be needed to test these possible explanations.

As suggested in previous psycholinguistic studies (Clark, 1990), there tends to be high cross-participant variation, but low within-participant variation, suggesting internal consistency. We find this in our own studies, as even across multiple trials in the same context and scenario, participants did not significantly change the number of marbles they selected. The implications of the higher variability of across participants suggest that the internal understanding of modal verbs in each individual differs, even one's own understanding of how to interpret the modals in a probabilistic context is consistent. It could be the case that when one is describing the probability of a more subjective event occurring, there could be discrepancies between the mutual understanding of the probability of the statement being true.

There are some limitations to the study given the online survey nature of the task. Although we kept attention check trials to ensure people paid attention to the task, there were some instances which made us question whether the instructions of the task were understood. For example, when analyzing the responses across trials, we noted that two of the subjects matched all of the modal verb expressions with 10 marbles, or 100% probability. Even though these

individuals passed the attention check trials, this result was not expected, and could have arose from a misunderstanding of directions for the online task. This limitation can be addressed in future studies by conducting in-person studies to ensure participants are properly following the task, and creating a version of the task that is not online.

In the current series of experiments we decided to focus on the use of affirmative statements. However, based on the low acceptance of all modal verbs in association with low probabilities, it could be the case that the negation of modal expressions (such as "might not") are more acceptable when referencing low probability events. Future work should investigate the use of negation of modal expressions when referencing low and high probability events. Furthermore, additional research can investigate changes in the use of modal expressions with greater distributions, since it is possible that speakers' states of higher-order uncertainty (i.e. subjective uncertainty in objective contexts) affect speakers' use of modal expressions in addition to the objective probability of the event (Herbstritt & Franke, 2017).

Overall, our findings from the study suggest that in an objective context of a binary distribution of marbles, people show a distinction between strong and weak modals. Epistemic modal verbs are often interpreted in both a logical and a pragmatic way, depending on the task. Furthermore, certain modal verbs, such as *can*, lend themselves to deontic interpretations naturally. The implications of this study deepen our understanding of modal verbs, given their complex linguistic nature. This can be useful in terms of how we teach non-native speakers modal verbs, given the limited range of modal verbs in the English language, the different uses in different contexts, and the lack of clear quantitative distinctions between the modal verbs make

them harder to understand and interpret. Additionally, it can also be beneficial in how children are taught probability, given the ranges of meaning each modal expression can have in quantification and the large usage of modal expressions in everyday life.

Future work will look at the development of epistemic modal verbs in children. Given our findings on the distinctions between weak and strong modals, we can study how children make these distinctions between modal verbs in a range of binary probability distributions. Children are exposed to formal probability around the age of 11 and 12 based on Common Core State Standards. This raises several interesting questions about the development of language and concepts. Do children need to obtain a formal understanding of probability before they begin to use modal expressions to represent probabilistic outcomes? Can linguistic representations of probability help children learn formal concepts related to uncertainty? Understanding the development of the semantic and pragmatic development of epistemic modality could further our understanding of the logical and pragmatic interpretations of epistemic modal verbs, as well as how it connects to their understanding of other mental concepts, such as probability.

References

Bates, Maechler, Bolker, & Walker. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Beyth-Marom, R. (1982). "How Probable is Probable: A Numerical Translation of Verbal Probability Expressions." Journal of Forecasting, 1, 257-269.

Celce-Murcia, M., Larsen-Freeman, D., & Williams, H. A. (1983). The grammar book: An esl/efl teacher's course.

Chapman, R. H. (1975). The development of children's understanding of proportions. *Child Development*, 46, 141–148.

Clark, D.A. (1990) Verbal uncertainty expressions: A critical review of two decades of research. *Current Psychology: Research and Reviews,* 9(3), 203–235

Coates, J. (1988). The acquisition of the meanings of modality in children aged eight and twelve. *Journal of Child Language*, 15(2), 425-434.

Frawley, W., Eschenroeder, E., Mills, S., & Nguyen, T. (2006). *The expression of modality*, (Vol. 1).

Grice, H. P. (1989). *Studies in the way of words.* Cambridge, MA: Harvard University Press.

Hahn U., Engelmann C. (2014) Grounding Epistemic Modality in Speakers' Judgments. In: Pham DN., Park SB. (eds) PRICAI 2014: Trends in Artificial Intelligence. PRICAI 2014. Lecture Notes in Computer Science, vol 8862. Springer, Cham

Herbstritt, M. & Franke, M. (2017). Proceedings of the 39th annual conference of the cognitive science society.

Hirst, W., & Weil, J. (1982). Acquisition of epistemic and deontic meaning of modals. *Journal of Child Language*, *9*(3), 659–666.

Holmes, J. (1982). Expressing doubt and certainty in english. *RELC Journal*, *13*(2), 9–28.

Horn Laurence, R. (1972). On the semantic properties of the logical operators in english. *University of California Los Angeles Ph. D. Dissertation Google Scholar*.

Jaime, Asuncion & Perez-Guillot, Cristina. (2015). A Comparison Analysis of Modal Auxiliary Verbs in Technical and General English. *Procedia - Social and Behavioral Sciences*. 212. 292-297.

Lassiter, D. & Goodman, N. (2015) How many kinds of reasoning? Inference, probability, and natural language semantics. *Cognition*, 126, 123-134

Lyons, John. 1977. *Semantics*, Vols. I and II. Cambridge: Cambridge University Press.

Mauboussin, A. & Mauboussin, M. (2018) If You Say Something Is "Likely," How Likely Do People Think It Is?. *Harvard Business Review*.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188.

Noveck, I. A., Ho, S., & Sera, M. (1996). Children's understanding of epistemic modals. *Journal of Child Language*, *23*(3), 621–643.

O'Grady, S., Starr, A., Griffiths, T. L., & Xu, F. (submitted). Non-symbolic probability judgments rely on integer approximation. *Journal of Experimental Psychology- Learning Memory and Cognition*.

O'Neill, D. K., & Atance, C. M. (2000). 'Maybe my daddy give me a big piano': The development of children's use of modals to express uncertainty. *First Language*, *20*(58), 029–52.

Ozturk, O., & Papafragou, A. (2015). The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development*, *11*(3), 191–214.

Papafragou, A. (2006). Epistemic modality and truth conditions. *Lingua*, 116, 1688–1702.

Papafragou, A. (2002) Modality and theory of mind: Perspectives from language development and autism. In: S. Barbiers, F. Beukema and W. van der Wurff (eds.), *Modality and its interaction with the verbal system.* Amsterdam: John Benjamins. 185-204.

Papafragou, A. (1998), The Acquisition of Modality: Implications for Theories of Semantic Representation. *Mind & Language*, 13: 370-399.

Piaget, J., & Inhelder, B., (1975). *The Origin of the Idea of Chance in Children*.

Reyna, V. F. (1981). The language of possibility and probability: Effects of negation on meaning. *Memory & Cognition*, *9*(6), 642–650.

Rubin, V. L. (2010). Epistemic modality: From uncertainty to certainty in the context of information seeking as interactions with texts. *Information Processing & Management*, *46*(5), 533–540.

Sweetser, E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge University Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive psychology, 5(2), 207-232.

Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). Two dimensions of subjective uncertainty: Clues from natural language.

Wagner, D., Dicks, J., & Kristmanson, P. (2015). Students' language repertoires for prediction. *The Mathematics Enthusiast*, *12*(1), 246.