Is There a "Voice" of Certainty? Speakers' Certainty Is Detected through Paralanguage

Alex B. Van Zant[1], Eduardo B. Andrade[2]

[1] Wharton School, University of Pennsylvania

[2] Brazilian School of Public and Business Administration, FGV, Rio de Janeiro

The expression of certainty is associated with a host of social and economic benefits, but we have no empirical knowledge of whether people are capable of using paralanguage (i.e., nonverbal cues like volume, pitch, and speech rate), to accurately infer others' certainty. Using an archival analysis of speakers making predictions about the likelihood of future events (Study 1) and two experiments that manipulate speakers' certainty in the accuracy of information (Studies 2 and 3), the current research demonstrates that listeners are capable of using speakers' paralanguage to make accurate inferences about their degree of probabilistic certainty. A meta-analysis of 1,577 recordings of 144 different speakers and 2,711 estimates made by 497 different listeners revealed that speakers and listeners alike relied on variance in pitch and speech rate to convey certainty and infer certainty in others, respectively.

Key words: certainty; paralanguage; vocal cues; confidence; perceptual accuracy

---

1. Introduction

Throughout his campaign that would eventually win him the 2008 United States Presidential Election, Barack Obama was noted for his speeches and sound bites that exuded a sense of certainty in his ability to deliver a future of "hope" and "change." Many have noted Obama's supreme confidence (Kantor 2008, Traynham 2012) and his success has prompted many communications experts to credit his strategic use of various vocal cues such as volume, variability, and pacing to convey a sense of unflappable certainty (Gallo 2012, Gentry 2009). Undoubtedly, one may be wise to employ these techniques, as the expression of certainty in the accuracy of one's judgment provides a variety of benefits including the attainment of social status (Anderson et al. 2012), the elicitation of others' trust (Sniezek and Van Swol 2001), the ability to influence others (Hinsz 1990, Sniezek and Van Swol 2001, Van Swol and Sniezek 2005, Zarnoth and Sniezek 1997) and the signaling of competence (Anderson et al. 2012). Clearly, these benefits have important implications for professional advisors and even organizational leaders. Advisors who convey certainty are more likely to be hired (Radzevick & Moore 2011) and less likely to have their advice challenged (Sah et al. 2013) than those who appear uncertain; leaders who express certainty can enhance their ability to inspire others to take action beneficial to their organizations (Flynn and Staw 2004, Westley and Mintzberg 1989) and are generally perceived as possessing characteristics that can improve organizational performance (Mendl et al. 1985).

Despite the well-known benefits of expressing certainty and claims that particular strategies may be effective in conveying certainty, it is an open question whether people are in fact capable of using paralanguage, or non-linguistic vocal cues (such as volume, pitch, and speech rate) to successfully signal their degree of *probabilistic certainty* (i.e., the probability they attach to their ability to forecast the future or provide accurate information) to others. The vast majority of research on the expression of probabilistic certainty has considered verbal displays of certainty through linguistic phrases (e.g., Tenney et al. 2007, Tenney et al. 2008) or numerical probability statements about the likelihood of a specific outcome (e.g., Price and Stone 2004, Radzevick and Moore 2011). However, paralanguage does not explicitly communicate a precise probability (cf. Budescu and Wallsten 1995), so it could easily fail to offer any informative value to listeners attempting to infer expressers' degree of probabilistic certainty.
Due to its potential to be misinterpreted by others, this suggests that speakers may not naturally rely on ambiguous paralanguage to convey certainty in favor of using less ambiguous verbal claims. Despite this sensible hypothesis, some indirect evidence suggests that speakers may actually rely more on paralanguage than verbal claims to convey certainty. Specifically, in a study by Anderson et al. (2012), the authors examine a variety of behaviors and find that among those behaviors, a "confident and factual vocal tone" had the strongest association with individuals' overconfidence on a weight-guessing task. Interestingly, this behavior was more strongly correlated with overconfidence than explicit statements of certainty made by participants in reference to the accuracy of their estimates. For paralanguage to be informative to listeners attempting to infer speakers' certainty, it is crucial that speakers use it to convey certainty in the first place. This indirect evidence raises the possibility that speakers may use it just as much, if not more, than explicit verbal claims to convey certainty.

So the question remains: Does speakers' paralanguage reveal information about their degree of certainty? That is, holding constant the verbal content of what speakers say, are listeners capable of using speakers' paralinguistic cues (e.g., volume, pitch, speech rate) to make accurate inferences about their degree of certainty? Answering this question is the primary goal of the current research. Additionally, should paralanguage be informative to listeners in interpreting a speaker's degree of probabilistic certainty, it is prescriptively informative to understand the specific cues used by speakers and listeners alike. Thus, a secondary goal of the current research is to examine whether particular cues are used by speakers to convey certainty and by listeners to infer certainty in speakers.

2. Conveying Certainty through Paralanguage

Scholars in the fields of linguistics and phonetics have been interested in paralanguage for quite some time. However, research on speakers' certainty displays is limited. One line of research concerns the use of vocal cues in revealing speakers' feeling-of-knowing (Brennan and Williams 1995, Kimble and Seidel 1991, Smith and Clark 1993). This work focuses on how individuals' confidence in their ability to retrieve information from memory influences their speech patterns. Given that cues indicative of memory search like speech latency (i.e., silences before speaking) and the use of linguistic fillers (e.g., "uh") correlate with speakers' self-reported feeling-of-knowing (Kimble and Seidel 1991, Smith and Clark 1993), this work suggests that paralanguage reveals information about speakers' self-perceived likelihood of having a specific piece of information stored in memory, but it does not speak to whether paralanguage reveals information about speakers' certainty in the accuracy of information once it has been retrieved.

Another line of research has, like us, focused on listeners' inferences about speakers' non-verbal paralinguistic cues (Aronovitch 1976, Kimble and Seidel 1991, Ramirez Verdugo 2005, Scherer et al. 1973). This work has been instrumental in providing evidence that listeners use paralanguage to draw inferences about speakers' general confidence, but it does not necessarily speak to whether this inferred confidence necessarily corresponds to an assumption that speakers believe there is a high probability of their judgment being accurate. In addition to certainty in the accuracy of one's judgment, confidence can also take the form of believing one's relative abilities are superior to others and believing one's task performance is high in an absolute sense (Moore and Healy 2008, Van Zant and Moore 2013). Thus, while listeners use paralanguage to infer that a speaker is generally confident in his or her task performance and relative abilities, they may not necessarily be able to use it to help them successfully differentiate between forecasts or information that the speaker is confident about from those that the speaker is less confident about. For example, when watching political candidates discuss the future state of the economy, people may spontaneously rely on the candidates' paralanguage to make inferences about how confident they are in their ability to competently perform the duties of their office in a particular economic climate, but they may not necessarily use it to make inferences about candidates' belief about the likelihood of a particular economic forecast proving to be accurate (or inaccurate). People often rely on others' judgment to help them differentiate between forecasts and intelligence that are likely to be accurate (i.e., a high probability of being true) from those that are not (Valley, White, Neale, &

Bazerman, 1992), so understanding whether paralanguage reveals anything about speakers' probabilistic certainty has value beyond simply knowing whether it reveals something about speakers' general dispositional confidence.

Moreover, and as important, there are methodological limitations to prior research that render it difficult to determine whether there actually is a causal link between a speaker's confidence and listeners' ability to accurately detect that confidence through the speaker's paralanguage. One limitation is that the majority of research only focuses on how paralanguage is either associated with a speaker's level of confidence (e.g., Kimble and Seidel 1991, Ramirez Verdugo 2005) or a listener's perception of the speaker's level of confidence (e.g., Aronovitch 1976) without actually accessing whether listeners are accurate at detecting speakers' true degree of confidence. Another limitation is that much of the research referenced above merely presents correlational evidence of a link between speakers' confidence and use of nonverbal cues without manipulating speakers' degree of confidence and fully controlling for the linguistic content of their speech (e.g., Aronovitch 1976, Ramirez Verdugo 2005). Speakers naturally vary their vocal cues when reading phrases that differ in their degree of linguistically implied confidence (Ramirez Verdugo 2005), so what may appear to be a causal relationship between speakers' confidence and their use of particular cues that could be used by listeners to infer their degree of certainty may merely be an artifact of differences in the verbal content of their speech when under certainty as opposed to uncertainty. Finally, research that actually does experimentally manipulate a speaker's confidence and assess listeners' judgments of the speaker's confidence (Scherer et al. 1973) is limited in generalizability due to the use of a single trained actor, who presumably possesses expertise at using paralanguage to convey emotional states and dispositions. This raises questions about the extent to which others share the actor's ability to convey confidence through paralanguage.

Though some researchers have asserted that paralanguage is more informative to interpreting others' degree of confidence than verbal statements (e.g., Maslow et al. 1971, Walker 1977), no research to our knowledge has provided causal evidence that people can use speakers' paralanguage to make accurate inferences about their probabilistic certainty—let alone whether lay speakers are even capable of readily transmitting this information through paralanguage. In this paper, we overcome this limitation by introducing a dynamic experimental paradigm whereby speakers are privately informed about the degree of certainty surrounding the veracity of a given statement and tasked with relaying this information to a paired listener without making linguistic alterations to the verbal content of statements. In other words, we asked speakers to communicate their objective degree of certainty and assessed whether listeners could accurately judge speakers' degree of certainty—*solely using paralanguage*.

3. Overview of Studies

Across three studies, we assess whether speakers' paralanguage reveals information about their degree of certainty. In a first study, we capitalize on an archive of recordings containing speakers' predictions of future event likelihoods and examine whether listeners are capable of accurately inferring their degree of probabilistic certainty—even when the recordings are devoid of any comprehensible linguistic content. In addition to identifying whether natural fluctuations in speakers' paralanguage provide meaningful information about their degree of certainty that can be perceived by others, this study also differentiates listeners' ability to infer speakers' degree of probabilistic certainty from their ability to infer speakers' degree of probabilistic doubt. Studies 2 and 3 address limitations to the first study by introducing an experimental paradigm that manipulates speakers' degree of certainty in the accuracy of statements while holding the linguistic features of the statements constant. The novel experimental paradigm uses an incentive-compatible compensation scheme to motivate speakers to convey a target level of certainty and listeners to accurately identify speakers' degree of certainty. Study 3 adapts the paradigm by removing speakers' financial incentive to convey a target level of certainty while allowing them to convey certainty in a more naturalistic fashion.

Should listeners be capable of accurately inferring speakers' degree of probabilistic certainty, then a secondary goal of the current research is to examine whether any particular set of cues are used by speakers to convey certainty and listeners to infer speakers' certainty. As such, we conclude with a meta-analysis of speakers' cues from the three studies.

4. Study 1

We start with an examination of whether speakers' naturalistic use of paralanguage reveals useful information to listeners attempting to infer speakers' degree of probabilistic certainty in the accuracy of their own forecast about the outcome of an event. To do this, we created an archive of recordings containing speakers who were tasked with estimating the likelihood of a future event occurring and justifying their estimates. Specifically, we recorded audio segments from the television show *Pardon the Interruption*, which airs on weekdays on ESPN. It is a sports talk show where two hosts discuss recent events in the world of sports. On occasion, the hosts participate in a segment called *Odds Makers* where they are presented with a series of future events and asked to give a prediction about the likelihood of each event occurring using a percentage. For each focal event, a moderator first presents the event before the two hosts take turns making their predictions and providing justifications for the predictions. After each prediction, the moderator records the host's numerical likelihood judgment on a chalkboard. A bell rings to signal to the hosts that they need to wrap up their discussion about each event so that the moderator can announce the next event they will evaluate. As an example, consider an event posed by the moderator on April 18, 2013 when he asked the hosts to estimate the chances that "Derek Jeter plays this season." One host spoke first for 35 seconds and estimated the likelihood of this event occurring to be 75%; after the moderator wrote down this prediction, the second host spoke for 29 seconds and estimated the likelihood of this event occurring to be 60%. After the moderator recorded the second host's prediction, a bell rang to signal a transition to the next event posed by the moderator. We recorded audio of all episodes featuring this segment beginning on April 9th, 2013 and ending on October 15, 2014 (the last known episode to include this segment as of November 1, 2015). For each prediction made by a given host, we created a recording that isolated the speech coming from his or her voice. While the hosts take turns in making their predictions during the segment, in some instances, a host's voice is briefly interrupted by the other host's voice, the voice of the segment's moderator, or a bell signaling the need for the hosts to wrap up discussion about the event. Because we are interested in isolating the focal speaker's paralanguage, we edited all recordings to cut out any sound coming from a source other than the focal speaker's voice. To do this, we cut out segments of each sound file as closely as possible to the start and end of any extraneous sound. In total, we have data on 228 predictions about 115 events that are made by eight different hosts across 30 episodes, and each host spoke on average for 24.58 seconds (*SD* = 5.69).[1]

After creating 228 recordings that each contain a single prediction made by a single host, we ran them through a low-pass filter. In line with prior research that has used a low-pass filter to destroy the intelligibility of speech while preserving elements of speakers' paralanguage, we removed all sound above 500 Hz and used 20 Hz smoothing around the cutoff (Scherer et al. 1984). To do this, we used the program Praat (Boersma and Weenik 2015). In addition to applying a low-pass filter, we also altered each sound file by standardizing the volume at 80 dB. We did this to ensure that the volume of each recording had a sufficiently high volume and to control for arbitrary differences in the volume of recordings, as they may vary slightly across episodes due to inconsistencies in the audio settings of the recorder at the studio where the show is taped.

This archive has several desirable features for the study of whether listeners can accurately infer speakers' probabilistic certainty using their paralanguage. First, for each event that speakers evaluated, they assigned a numerical probability to its likelihood. This enables a normative comparison between listeners' estimates of speakers' certainty and speakers' actual certainty that would not be possible with a more subjective measure. By using audio-filtering techniques to render recordings in the archive incomprehensible from a linguistic standpoint while still preserving properties of speakers' paralanguage, we were able to present research participants with modified recordings from the archive while asking them to predict speakers' objective degree of probabilistic certainty without any knowledge of the event

[1] Though each of the two hosts are meant to provide a prediction of each event, there were two instances where one host did not make a prediction in favor of deferring discussion about the event to the other host. As such, we have recordings of 228 total predictions rather than 230.

speakers were evaluating. Thus, we could assess whether listeners are accurate at inferring speakers' degree of probabilistic certainty when they only have access to an audio file containing traces of speakers' paralanguage.

A second desirable feature of this archive is that speakers engage in a naturalistic discussion of the reasoning behind their forecasts. While much prior research on the vocal cues associated with confidence uses brief statements (e.g., Brennan and Williams 1995, Smith and Clark 1993) or single utterances (e.g., Kimble and Seidel 1991), a strength of this archive is that it allows us to assess whether research participants can still determine speakers' degree of probabilistic certainty when the portion of their speech where they actually provide a numerical forecast is buried in a longer segment of speech. In this manner, we assess whether listeners can detect speakers' degree of certainty even when they are not necessarily cued to the specific portion of speech where speakers reveal their degree of certainty—a feature that enables a conservative test of perceptual accuracy.

Another beneficial feature of this archive is that speakers provided estimates of event likelihoods using a full scale ranging from 0% to 100%, which allows us to explore whether listeners' ability to detect speakers' level of probabilistic certainty (0% or 100% outcome likelihood) vs. uncertainty (50%) is distinguishable from their ability to detect speakers' level of probabilistic doubt (0%) vs. certainty (100%). Because a speaker's assignment of 50% to an event's likelihood can either be interpreted as complete uncertainty about whether the event will occur or as moderate doubt about whether the event will occur (as compared to complete doubt at 0%), this distinction is valuable for understanding whether perceivers' ability to detect speakers' degree of certainty is distinct from their ability to detect the extent to which speakers doubt that an event will occur. To explore this distinction, we randomly assigned some participants to estimate speakers' predictions on an uncertainty-certainty dimension and others to estimate the same predictions on a doubt-certainty dimension.

3.1. Participants

Though we did not have any preconceived notions about the effect size of detection accuracy in either experimental condition, we wanted to collect a sample size large enough to detect two effects with at least 95% power: A main effect of detection accuracy in the uncertainty-certainty condition greater than or equal to $r = 0.1$ and an interaction effect of the same magnitude should the effect of detection accuracy differ across conditions. Using the G*Power software (Faul et al. 2007, 2009), we determined that this would require a minimum of 2,578 evaluations spread across at least 652 different research participants (assuming no correlation between a given participant's evaluations). To achieve a sample size of at least this magnitude, we recruited 1,079 Amazon Mechanical Turk workers to complete a study on communication in exchange for a minimum payment of $0.75. Given that it was critical for participants to manually play recordings, to have their speakers on at a reasonable volume, and to understand the nature of the task, we included two multiple-choice attention checks. One of them was a sound check that required participants to accurately report information contained in a recording that they played on the prior page while the other was a comprehension check that required them to accurately identify the task they would be performing in the study. We ultimately collected 2,747 evaluations from 728 participants who passed both checks (67%); their mean age was 34.41 years (*SD* = 11.75) and 45% of them were female. Those who did not pass both comprehension checks were not allowed to continue with the study and we therefore did not collect any measures from them.

3.3. Procedure

At the start of the study, all participants were informed that they would listen to a series of recordings where they would not be able to "comprehend what the speaker is saying." To better clarify this statement, we told participants the following: "Though you will not be able to hear what the speaker says, you will be able to hear the way that he or she says it." We then briefly summarized the premise of the *Odds Makers* segment on *Pardon the Interruption*. To help them better understand the context surrounding the recordings they would later evaluate, we presented participants with a sample recording similar to the ones they would later be hearing, with the exception that it had yet to be filtered and was therefore comprehensible. Participants then answered a comprehension check question asking them to recall the likelihood assigned by the speaker to the event discussed in the recording. At this point, they received directions explaining that they would hear similar clips that had been "filtered in a way that makes the speaker's words incomprehensible." They then answered a comprehension question about the procedure before being presented with an incomprehensible filtered version of the sample recording they previously heard. As the goal was to give participants a sense of the procedure before beginning the actual study, we included a scale for eliciting participants' estimates of the speaker's level of uncertainty-certainty or doubt-certainty below the box where the recording could be accessed. This presentation was identical to what participants saw during the actual judgment task.

Participants then proceeded to make their own estimates of speakers' uncertainty-certainty or doubt-certainty. We randomly assigned participants to hear a block of filtered recordings containing all of a given speaker's predictions from a given episode, which varied from 2-4 recordings. As there are 60 different speaker-episode combinations in the archive of recordings, participants were randomly assigned to one of 60 speaker-episode blocks. Within each block, recordings were presented in a sequential order such that participants heard a recording and made an estimate about the speaker's uncertainty-certainty or doubt-certainty before proceeding to the next recording. To incentivize accuracy during the task, we offered a $25 bonus prize to the participant whose estimates came closest to speakers' actual predictions.

Participants were randomly assigned to one of two conditions: An uncertainty-certainty condition where they estimated speakers' certainty in the accuracy of their binary belief about whether an event would occur or not and a doubt-certainty condition where they simply estimated speakers' likelihood estimates for each event.

3.3.1. Uncertainty-certainty condition. In the uncertainty-certainty condition, participants estimated speakers' certainty in their predictions about whether an event would occur or not (i.e., the extremity of speakers' predictions relative to 50%). At the start of the experiment, participants assigned to this condition read that they would "listen to several recordings containing a speaker's explanation of how certain he or she is about a series of predictions." After the first comprehension check, we then explained what this means using the sample clip as an example. Specifically, we told participants the following:

> In the recording you just listened to, the speaker was discussing an event that he perceives as having a 40% likelihood of happening. In other words, the speaker predicted that the event was most likely *not* going to happen. However, because the speaker believed there was still a 40% chance of his prediction not being accurate (i.e., the probability assigned to the opposite prediction), he was not very certain about his prediction. In other words, the speaker only believed that the likelihood of his prediction being accurate was 10% above a chance rate, or 60%. Had the speaker provided a likelihood further away from 50%, the speaker would have been more certain about his prediction. An estimate of 50% would have indicated that the speaker was completely uncertain about his prediction.
>
> In this study, you will listen to similar clips of a speaker making predictions and indicating how certain he or she is about each prediction.

Thus, we framed participants' task in this condition as being to estimate speakers' degree of certainty in a prediction about whether an event would occur or not. When they attempted to estimate speakers' certainty, they did so by filling in the blank of the following statement for each recording: "The speaker thinks there is a ___% chance of his or her prediction being accurate." They filled in the blank by selecting number on a sliding scale ranging from 50 (*completely uncertain*) to 100 (*completely certain*).

3.3.2. Doubt-certainty condition. In the certainty-doubt condition, we wanted participants to estimate speakers' likelihood judgments in a straightforward manner. At the start of the experiment, participants assigned to this condition read that they would "listen to several recordings containing a speaker's explanation of how likely he or she thinks a series of events are to happen." After the first comprehension check, we explained what this means using the sample clip as an example:

> In the recording you just listened to, the speaker was discussing an event that he perceives as having a 40% likelihood of happening. In this study, you will listen to similar clips of a speaker discussing

different events and assigning numbers to how likely the speaker thinks it is that each event happens in the future.

Thus, we framed participants' task in this condition as being to estimate speakers' perception of how likely a focal event is to occur in the future. When they attempted to estimate speakers' beliefs, they did so by filling in the blank of the following statement for each recording: "The speaker thinks there is a ___% chance of the event happening." They filled in the blank by selecting number on a sliding scale ranging from 0 (*no chance of happening*) to 100 (*certain to happen*).

3.3.3. Comprehension checks. Although the instructions and procedures varied slightly by experimental condition, an examination of the comprehension checks showed that they had no impact on participants' likelihood of proceeding to the actual judgment task by correctly passing both checks, $\chi^2(1, N = 1079) = 0.13$, $p = .72$. Thus, any observed variance across conditions is unlikely to be attributable to differences in participants' attentiveness or understanding of the procedures.

3.3.4. Speaker forecasts. We created condition-specific measures of speaker forecasts to use as a comparison point for participant estimates. This was straightforward in the doubt-certainty condition: We simply used speakers' likelihood estimates on each recording (*Likelihood*). However, in the uncertainty-certainty condition, we needed to transform this variable to capture the extremity of speakers' forecasts relative to 50%, or the point at which they perceived an event as equally likely to occur as to not occur. Because the purpose of this dimension is to capture speakers' certainty in the accuracy of their binary prediction about whether an event will occur or not independently of the actual prediction itself, we used the absolute value of the difference between speakers' likelihood estimates and 50. To correct for this transformation's range reduction (i.e., the maximum value becomes 50 rather than 100), we multiplied the resulting value by 2. Thus, in the uncertainty-certainty condition, the speaker forecast measure was:

$$2 * |Likelihood - 50|$$

3.3.4. Participant estimates. The key dependent measure was participants' estimates of speakers' forecast as indicated on the appropriate condition-specific scale. This was simply the percentage that a participant selected for a given recording in the doubt-certainty condition. Because participants in the uncertainty-certainty condition estimated speaker forecasts using a scale that captured speakers' certainty in the accuracy of their predictions, we only needed to adjust the dependent measure in this condition to account for the scale's restricted range relative to the doubt-certainty condition (participants in the uncertainty-certainty condition used a scale ranging from 50% to 100% rather than 0% to 100%). To do this, we subtracted 50 from the percentage that a participant selected for a given recording and then multiplied this result by two.

3.4. Results

We tested for a detection accuracy effect by regressing participant estimates on speaker forecasts. Our models account for the hierarchical nature of the dataset using a participant-specific random intercept. We also use fixed effects to control for idiosyncratic speaker characteristics, episode fixed effects to account for the possibility of arbitrary differences in studio recorder settings, and the order in which participants evaluated recordings. As documented in Table 1 (Model 1), we found that participant estimates predicted speaker forecasts, ($z = 4.80$, $p < .001$, $\beta = .10$). That is, despite not being capable of understanding the verbal content of speakers' recordings and having no information to rely on aside from speakers' paralanguage, participants were somewhat capable of accurately identifying speakers' forecasts. This effect was qualified by an interaction with the evaluative dimension treatment ($z = 2.35$, $p = .019$). Participants were more accurate at predicting speakers' forecasts in the uncertainty-certainty condition (Model 3: $z = 4.41$, $p < .001$, $\beta = .15$) than they were in the doubt-certainty condition (Model 4: $z = 2.29$, $p = .022$, $\beta = .06$). As detailed in the SOM, the main effect of detection accuracy in the uncertainty-certainty condition (Table S2) and the interaction effect with the evaluative dimension treatment (Table S1) are robust across alternative model specifications.

Though participants in the doubt-certainty condition were not quite as accurate at predicting speakers' forecasts as those in the uncertainty-certainty condition, they still displayed some degree of accuracy. This presumably indicates that they could detect speakers' belief about whether an event would happen or not (i.e., whether their forecast for an event's likelihood was closer to 0% or closer to 100%). We found evidence consistent with this in a set of post-hoc exploratory analyses (detailed in the SOM), as participants in this condition could detect whether speakers doubted that the event would occur (i.e., a speaker perceived an event has having a sub-50% likelihood of occurrence) at a rate above chance, $B = .51$ ($SE = .13$), $z = 3.87$, $p < .001$, $OR = 1.67$. However, they were not only accurate at detecting speakers' directional belief in this condition. Consistent with the findings in the uncertainty-certainty condition, they also displayed some accuracy at detecting speakers' probabilistic certainty, $B = .06$ ($SE = .03$), $z = 2.20$, $p = .032$, $\beta = .07$. Thus, participants could on some level detect two classes of paralanguage: One that reveals information about speakers' directional belief about whether an event will occur and one that reveals information about speakers' certainty in that belief. However, as we document in the SOM (Section 1.2), their reliance on speakers' certainty undermined their ability to accurately identify speakers' directional belief. This not only further documents that perceivers can use speakers' paralanguage to detect their degree of probabilistic certainty, but it also provides evidence that perceivers' ability to accurately detect speakers' forecasts on an uncertainty-certainty dimension is independent of speakers' use of paralanguage to signal whether they doubt that an event will occur.

Table 1. Study 1: Regressions Predicting Participant Estimates

	Both Conditions		Certainty-Uncertainty Condition	Certainty-Doubt Condition
	Model 1	Model 2	Model 3	Model 4
Speaker Forecast	0.07 (0.01)***	0.04 (0.02)	0.11 (0.02)***	0.05 (0.02)*
Uncertainty-Certainty	-6.84 (0.98)***	-10.51 (1.85)***		
Speaker ForecastXUncertainty-Certainty		0.07 (0.03)*		
Speaker Fixed Effects	Yes	Yes	Yes	Yes
Episode Fixed Effects	Yes	Yes	Yes	Yes
Statement Order Fixed Effects	Yes	Yes	Yes	Yes
N	2,747	2,747	1,361	1,386
Log Likelihood	-12,565.11	-12,562.37	-6,261.75	-6,261.29

Note. Numbers represent coefficient estimates (standard errors in parentheses). All parameters were estimated using linear regression models with a participant-specific random intercept that was estimated using maximum likelihood estimation. *Uncertainty-Certainty* = 1 for participants in certainty-uncertainty condition, 0 for participants in certainty-doubt condition.
$^{*}p < .05$. $^{***}p < .001$.

3.5. Discussion

Using recordings of extended naturalistic speech that had been edited to contain traces of speakers' paralanguage while being linguistically unintelligible, we found evidence of an effect where listeners could accurately detect speakers' certainty in the accuracy of their predictions about future events. That is, listeners could infer speakers' degree of certainty based simply on the *way* speakers were talking rather than on *what* speakers were actually saying. Furthermore, because participants were more accurate when estimating speakers' certainty in the accuracy of their predictions than they were at estimating the probabilities that speakers assigned to the likelihood of events, we suggest that paralanguage reveals information about speakers' degree of probabilistic certainty (vs. uncertainty) that is distinct from and more readily decoded than any information it reveals about speakers' degree of probabilistic doubt in an event's likelihood.

While its findings are consistent with the notion that speakers' paralanguage reveals information about their degree of certainty that is readily encoded by listeners, Study 1 has three key limitations. First, the sample of speakers is limited in quantity and represents media pundits, who may have specialized training encouraging them to encode paralanguage in a manner that is readily interpreted by others. Second, because we measure speakers' degree of certainty using their own self-reports, we cannot rule out the possibility of reverse-causality where speakers' paralanguage over the course of speaking influences their self-reports (in contrast to our assumption that differences in speakers' paralanguage are caused by their degree of certainty). Finally, the modifications we made to speakers' recordings limits our ability to make complete inferences about the cues that listeners use to infer speakers' certainty. Though our technique of using a low-pass filter to destroy speech intelligibility retains some properties of speakers' paralanguage, it fails to retain many other properties that are valuable to making accurate inferences about speakers' mental state (Scherer 1971, Scherer et al. 1984). Furthermore, despite the presence of prior research documenting that speech volume is associated with speakers' certainty (Aronovitch 1976, Kimble and Seidel 1991, Scherer et al. 1973), we standardized the volume of all recordings presented to participants, which effectively made a likely cue to speakers' certainty inaccessible to perceivers. Taken together, these modifications to recordings render Study 1 a conservative test of whether listeners can use paralanguage to infer speakers' degree of certainty, but it also limits our ability to understand the cues they use to do so.

4. Study 2

To address limitations to Study 1, we designed an experiment that allowed us to manipulate lay speakers' degree of certainty while providing perceivers with unedited samples of speech that fully retained speakers' use of paralanguage, yet did not reveal any linguistic information about their degree of certainty. In the experiment, we presented speakers with a set of statements and private knowledge about whether each statement *is true* (100% likely to be accurate) or *may be true* (50% likely to be accurate). In this way, we manipulated speakers' degree of certainty in the accuracy of information. Speakers were then tasked with audio-recording the statements in a way that could allow an unknown listener with whom they have been paired to accurately estimate the respective degree of certainty associated with each statement; we prohibited them from making alterations to the verbal content of statements and monitored for failures to follow this instruction. Listeners then played the audio-recorded statements and indicated whether the speaker believed each statement to be true or possibly true. To incentivize accuracy, both the speaker and his or her paired listener earned $0.50 for each accurate estimate.

3.1. Pretesting Statements

To identify an appropriate sample of statements for the experiment, we first generated forty sentences representing factual statements (e.g., "The waterproof wristwatch was invented by Rolex in 1923."). For each factual statement, we generated an analogous, but false, version of the same sentence (e.g., "The waterproof wristwatch was invented by Rolex in 1933."). The false version of each sentence was written in a manner that is linguistically similar to the true version. To generate factual statements, we consulted a variety of online resources. As of March 16, 2012, the factual versions of each statement were considered to be true.

We then pre-tested these statements for the purposes of identifying those that research participants had no preconceived notions about whether they were actually true or false. This ensured that we could implement a deception-free manipulation of the degree to which participants should be certain about the veracity of the statements they read. A pretesting sample of forty-three undergraduate business students read either the factual or inaccurate version of each statement in exchange for course credit. For each statement that participants read, either the true or the false version was randomly presented. After reading each statement, we asked participants to indicate both the likelihood of the statement being true and the likelihood of the statement being false, conditional on both probabilities adding to 100%. With the goal of being able to manipulate participants' degree of certainty in the veracity of statements for the main study without using deception, we sought to identify sets of statements meeting two key criteria. The first criterion was that participants did not consider the true or the false version of a given statement to have a likelihood of being true that significantly differed from 50%. The second criterion was that for both versions of a given statement, participants' estimate of the statement's likelihood of being true should not significantly differ from their estimate of the statement's likelihood of being false. We identified twelve statements meeting this set of criteria and incorporated them into a computer program created for the experiment.

3.2. Participants

Forty undergraduate business students enrolled in an undergraduate business course participated in the experiment in exchange for course credit and the opportunity to earn up to $6 in performance incentives. We selected this sample size because it was the maximum number of participants we could recruit from that particular class. Participants who engaged in the pretesting portion of this study were excluded from participating in the main experiment. Participants' mean age was 21.89 years (SD = 3.11 years). Ten participants were male (25%).

3.3. Procedure

After identifying suitable statements and developing a computer program for administering the experiment, we conducted a laboratory experiment where participants were seated in private rooms, randomly paired with another participant in the same experimental session, and incentivized to strategically convey certainty in one role (*speaker*) as well as to estimate/guess another's degree of certainty in another role (*listener*). Within each set of paired participants, we randomly assigned one member of the pair to a set of six statements and the other to a different set of six statements. For each statement that participants read, we randomized the information they had about the statement's likelihood of being true.

We recruited participants in groups of 2-4 for each experimental session. An experimenter led each participant into his or her own private room immediately upon entering the laboratory to prevent participants from interacting with one another prior to the start of the experiment. Seating participants in private rooms immediately upon their entry into the laboratory also ensured that they would be unaware of the number of participants in their session. The experimenter seated each participant in front of a Lenovo IdeaPad U350 laptop with a pre-installed Microsoft high definition microphone. Noise-cancelling headphones were plugged into each laptop. All laptops were identical and the microphone settings were at the same levels for all experimental sessions. Once an experimental session was underway, the experimenter started the computer program for each participant and ensured that all participants were wearing a pair of noise-cancelling headphones. The laptop speakers were set to the maximum possible volume to maximize the audibility of recordings played through the headphones.

Participants opened the computer program to start an experiment on vocal communication. Upon starting the program, they read instructions specifying that the "amount of money you earn will depend partly upon your own ability to nonverbally convey information and on another participant's ability to detect this information in your voice." Participants then advanced to a screen informing them that they would occupy two roles during the course of the experiment: message senders (*speakers*) and message receivers (*listeners*). As speakers, they were tasked with recording statements that were certainly true or possibly true. Their goal was to read each statement in a way that would allow another participant who they were randomly partnered with to guess whether the statement was certainly true or possibly true. They were required to read statements verbatim and forbidden from altering the verbal content of statements in any way. As listeners, they listened to recordings from their partner and guessed whether each statement was possibly true or certainly true.[2]

Participants always played the speaker role first and the listener role second. Thus, they recorded all of their statements before being exposed to any statements recorded by their randomly assigned partner. By playing the speaker role prior to the listener role, participants had yet to be exposed to any stimuli from their partners, thus making their strategic use of paralinguistic cues independent of any cues that may have been utilized by their partner. Prior to making their own voice recordings, participants completed a training session teaching them to generate recordings, play back their own recordings, and

---

[2] Because we were unsure whether we could adequately measure participants' vocal cue usage prior to starting data collection, we also asked speakers to indicate the extent to which they attempted to use a variety of vocal cues and convey particular emotions after recording each statement. Additionally, listeners indicated the extent to which they heard the same cues and emotions in each recording by replaying them. However, they were not able to replay recordings for the purposes of filling out these measures until they had already estimated speakers' degree of certainty for all six recordings. In light of the fact that (1) we were capable of obtaining objective measures of vocal cue usage and (2) speakers' self-reported use of emotions was near the scale floor for all emotions, we did not analyze these variables.

send recordings to their partner. Participants had the opportunity to generate up to three recordings of each statement in order to ensure that they were satisfied with the volume and general intonation of the recordings that were sent to their assigned listener. Though the program allowed them to play back their own recordings and to make multiple recordings of a given statement, they could only send their most recent recording of the statement. After completing the tutorial, participants went on to record a randomly assigned set of six statements in the speaker role. Once both members of a given pair of participants had finished recording statements in the speaker role, they advanced to the listener role. First, they completed a tutorial that taught them to listen to their partner's recordings using a pre-recorded statement prepared specifically for the tutorial. Once in the listener role, participants listened to a set of six statements recorded by their partner for a maximum of three times apiece (to ensure that they adequately heard the statements they were judging). The statements recorded by their partner were different from the statements that they recorded themselves. For each statement that their partner correctly identified as certainly true or possibly true, the speaker and listener earned $0.50 apiece. In other words, participants' incentives were aligned and they were fully aware of the rules of the task.

After completing the listener role, participants answered a brief demographic questionnaire. Once both members of a given pair of participants had completed the listener role, the computer program computed the pair's payoff and an experimenter paid participants as they completed the demographic questionnaire.

3.3.1. Statement set manipulation. We generated two different sets of six statements (see Appendix A for a complete list of statements). One member of each dyad was randomly assigned to record themselves reading Statement Set A while the other was assigned to Statement Set B. Statements within each set were presented to participants in a randomized order. Participants could only view and record one statement at a time in a sequential order.

3.3.2. Statement certainty manipulation. We provided direct information to speakers about the likelihood of a given statement being true. Speakers then read instructions specifying that each statement they read in the experiment would randomly come from one of two buckets: a bucket comprised of statements that *may be true* (uncertain statements) and a bucket comprised of statements that *are true* (certain statements). The instructions specified that statements that *may be true* had a 50% probability of being true while statements that *are true* had a 100% probability of being true.

When speakers read a statement from the bucket of statements that *are true*, they always saw the true version of the statement. When speakers read a statement coming from the bucket of statements that *may be true*, they were randomly presented with either the true version or the false version of the statement. The computer program clarified this distinction for speakers. For each individual statement, the computer program randomly determined whether it came from the bucket of statements that *may be true* or from the bucket of statements that *are true* with the constraint that a minimum of one statement come from each bucket.

3.3.3. Listeners' perceptions of speakers' certainty. Our primary dependent measure of interest was whether a listener considered a given message to have been drawn from the bucket of statements that *may be true* or from the bucket of statements that *are true*. Listeners indicated their judgment for each statement by either clicking a box reading "The Sender [speaker] wants to convey that this statement IS TRUE" or clicking a box reading "The Sender wants to convey that this statement MAY BE TRUE". Listeners could only listen to and form a judgment on one statement at a time in a sequential order. We note that this measure is a binary measure of speaker certainty, unlike the continuous metric we used in Study 1. This binary measure not only allows us to use an easy-to-interpret metric to summarize the extent to which listeners were accurate at identifying speakers' certainty (i.e., the percentage of the time they made an accurate categorization), but it also is a more appropriate match for the manner in which we manipulated speakers' certainty.

3.4. Results

Three statements were not recorded due to participant error, so our final dataset includes 237 statements recorded by all 40 speakers. Our behavioral data show that despite having no ability to display visual cues, leeway to verbally alter their messages, or knowledge about the identity of their paired listener, listeners were capable of successfully inferring speakers' degree of certainty at a 79% clip, a rate well above chance ($p < .001$). To more formally test for an accuracy effect, we ran a logistic regression using speakers' certainty to predict listeners' estimates of their certainty. Because participants were paired in dyads, we used a dyad-specific random intercept (estimated using adaptive quadrature with 30 integration points) to account for the hierarchical nature of the dataset, in addition to fixed effects to control for the order in which statements were evaluated. This model confirmed that listeners could detect speakers' certainty at a rate above chance, $B = 2.89$ ($SE = .35$), $z = 8.06$, $p < .001$, $OR = 18.05$; this effect held using alternative model specifications (see Table S4 in SOM). We also conducted a check to ensure that the verbal content of speakers' recordings did not deviate from the statements they were presented with. Because we failed to identify any such recordings, listeners' high accuracy cannot be explained by speakers ignoring experimental instructions by providing explicit information about their degree of certainty. Overall, this finding provides causal support for the notion that lay speakers are capable of using paralanguage in a manner that can allow listeners to accurately infer their degree of certainty.

3.5. Discussion

The results of Study 2 support the notion that lay speakers use paralanguage in a manner that allows listeners to readily infer their degree of certainty. Though promising, this finding should be interpreted with caution. Despite our care in ensuring that experimental manipulations were unlikely to have systematically influenced listeners' judgments or speakers' cue usage on any dimension other than certainty per se, we suspect that our effects may be larger in magnitude than one might expect to see in many other settings.

Two methodological features account for this suspicion. First, participants were strictly prohibited from providing any verbal information about their true degree of certainty. While this instruction was essential to our ability to infer that individuals can accurately detect others' degree of certainty solely via the use of paralanguage, it placed speakers in the unusual position of not being allowed to articulate their degree of certainty via verbal channels. This may have forced them to rely on modulating their vocal cues to a greater extent than they would in normal conversation where they can provide explicit linguistic information about their degree of certainty. Evidence suggests that individuals do not rely heavily on explicit linguistic statements of certainty when expressing their belief that they know the truth (Anderson et al. 2012), but it could be possible that constraining their ability to qualify statements with verbalized information led to exaggerated paralanguage. Out of concern that listeners would not be able to adequately infer their degree of certainty without access to relevant verbal information, speakers may have expressed certainty and uncertainty using a caricature of their normal speech patterns.

The second methodological feature of this study that may limit its generalizability is that we incentivized speakers to encode certainty in a manner that could be interpreted by listeners. While people may possess a fundamental motive to be perceived accurately by others (Human and Biesanz 2013), providing economic incentives to do so may elevate their motive beyond what may exist in many other contexts without such incentives. In providing speakers with some incentive to give the task a genuine effort, we may have inadvertently elicited an unrealistic degree of effort that caused speakers to deliberately manipulate their speech patterns in a manner that exceeds what they would do in typical conversation.

4. Study 3

In light of limitations to Study 2, we sought to replicate the accuracy effect in Study 3 while adapting the experimental paradigm to provide for a more conservative test of the effect. We had two primary goals in this experiment: 1) allow speakers an opportunity to qualify their statements with explicit verbal information about their degree of certainty and 2) remove accuracy incentives that may have encouraged speakers to exaggerate their paralanguage in order to be accurately perceived by listeners.

4.1. Participants

In this study, we recruited speakers and listeners from separate participant pools to complete the study at different points in time. Speakers were ninety-nine undergraduate business students who participated in exchange for course credit. Their mean age was 21.68 years (*SD* = 3.20) and 51 of them were female (52%). Listeners were ninety-nine participants recruited from Amazon Mechanical Turk who completed the study in exchange for $2.25. Their mean age was 30.65 years (*SD* = 8.13) and 39 of them were female (39%). To ensure that listeners understood the task, we required that they pass a series of reading and listening comprehension checks before proceeding with the study. Only six listeners failed to pass all of the comprehension checks. They were prevented from continuing with the study before we could collect data from them. Whereas listeners had a chance to earn a $1 bonus for accuracy, speakers were not provided with any such incentives.

4.2. Procedure

Our experimental procedure differed between speakers and listeners. In contrast to Study 2, speakers did not play the listener role and listeners did not play the speaker role.

4.2.1. Speakers. We recruited participants to play the speaker role in groups of 3-5. They were seated in their own private room in front of a laptop identical to the one participants used in Study 2. Due to concerns that the volume of recordings was insufficient in Study 2, we equipped participants with clip-on microphones connected to their laptop and adjusted the microphone volume to maximal levels on all laptops. Other than this adjustment, all laptop settings were identical to Study 2.

Speakers first read a set of general instructions informing them that they would be asked to read a series of statements. We then informed them that we wanted them to “have an opportunity to communicate this information in a manner that is similar to how you would do so on a day-to-day basis.” Speakers then proceeded to select specific phrases that they typically use to convey certainty and uncertainty. They first selected a phrase that they would use to communicate that something has a 100% probability of being true before selecting a phrase that they would use to communicate that something has a 50% probability of being true. See Appendix B for participants’ instructions and the list of phrases from which they chose.

Following the selection of preferred probability phrases, they played the speaker role as in Study 2, where each statement they read was randomly drawn from either the bucket of statements that *are true* or the bucket of statements that *may be true*. However, there were three key differences. First, because speakers did not later play the listener role, they read the all twelve statements from Study 2 as opposed to being assigned to a subset of six statements. Second, we modified the statements to include qualifying information about speakers' degree of certainty in their likelihood of being true. Specifically, we prefaced each statement with either the probabilistic phrase that a given speaker selected to convey certainty (for statements that *are true*) or uncertainty (for statements that *may be true*). For example, a participant who chose the phrase "I am certain that" to convey certainty would have read the phrase "I am certain that James Monroe was 73 years old when he died" drawn from the bucket of statements that *is true*. In contrast, the same participant may have chosen the phrase "I am uncertain whether" to convey uncertainty and would have read the same version of the statement as "I am uncertain whether James Monroe was 73 years old when he died" drawn from the bucket of statements that *may be true*. The final difference from Study 2 is that speakers were not provided with any economic incentives to have listeners accurately detect their degree of certainty. Instead, they were simply instructed to "read each statement in a way that could allow another person to guess whether the statement you read ***is*** vs. ***may be*** true just by listening to your recording."

3.2.2. Listeners. Before recruiting participants to play the listener role, we modified speakers' recorded statements so that they were linguistically identical to those in Study 2.[3] To do this, we eliminated the explicit verbal probabilistic information that participants used to preface their statements. For example, the phrases "I am certain that James Monroe was 73 years old when he died" and "I am uncertain whether James Monroe was 73 years old when he did" were both modified by cutting as closely as possible to the spoken potions of utterances "I am certain that" and "I am uncertain whether" so that listeners would hear "James Monroe was 73 years old when he died" for both versions of the statement. After completing this process for all participants' recorded statements, we uploaded them into a survey and recruited listeners to participate in the study.

Listeners read a series of instructions adapted from the listener role in Study 2 to explain the context of the experiment used to elicit speakers' recordings. They were randomly assigned to all of the statements recorded by a single speaker (presented in the same order that the speaker recorded them) and asked to "guess whether the speaker was trying to convey that each statement IS TRUE or MAY BE TRUE." To incentivize performance, we informed listeners that they would have a chance to earn a $1 bonus payment for making correct guesses on all of their statements. As in Study 2, listeners were presented with a single statement at a time and could not advance to the next statement until they

[3] To prevent participants from deviating from their natural speech patterns, we did not inform them about this planned modification to their recordings.

provided a guess. To prevent participants from looking up the veracity of statements online, we allotted them a maximum of thirty seconds to listen to each statement and make a guess before automatically changing the page to the next statement.

3.2.3. Listeners' perceptions of speakers' certainty. As in Study 2, our primary dependent measure was listeners' perception of speakers' degree of certainty on each recording.

4.3. Results

Though participants were collectively presented with 1,188 statements to record themselves reading, participants made errors on a total of thirty-nine statements in which they failed to record themselves. This left a final sample of 1,149 audible recordings that we presented to listeners. Though they did so to a lesser extent than participants in Study 2, listeners accurately inferred speakers' degree of certainty on 55% of recordings, a percentage greater than what would be expected by chance ($p < .001$). We confirmed this effect using a logistic regression model with a random intercept to represent speaker-listener dyads (estimated using adaptive quadrature with 30 integration points) and statement order fixed effects, $B = .69$ ($SE = .13$), $z = 5.17$, $p < .001$, $OR = 2.00$.[4]

4.4. Discussion

Though smaller in magnitude, we replicated the accuracy effect documented in Study 2. Unsurprisingly, our efforts to provide for a more conservative test of the effect resulted in listeners being less accurate at detecting speakers' degree of certainty in comparison to Study 2. Nonetheless, the results provide additional causal evidence that even in the absence of explicit verbal statements of certainty, listeners are

---

[4] After completing data collection, we identified a glitch in the randomization process determining whether a given statement was true vs. may be true. Specifically, only 38% of statements were drawn from the bucket of statements that were true. We conducted a series of checks to ensure that this did not influence the listener accuracy effect. First, we found that listeners' guesses were biased in the opposite direction such that they were more likely to consider a statement true as opposed to possibly true (58% vs. 42%). This renders the possibility that a directional bias in listener guesses may have interacted with the randomization bias to produce an overall accuracy effect unlikely. As a second check, we tested whether the randomization bias was uniform across statements. Though the bias did not vary by statement order, it did vary by statement type; Wald test: $\chi^2(11, N = 1{,}149) = 35.65$, $p < .001$. A follow-up investigation revealed that there was no glitch for three statement types that had an equal probability of being drawn from either bucket (49% of statements were drawn from the *is true* bucket, which did *not* differ from chance, $p = .77$) while the remainder of statements were biased to a similar degree (35% of statements were drawn from the *is true* bucket, which did differ from chance, $p < .001$). To ensure that the accuracy effect was not driven by any particular subset of statements, we replicated our analyses on both the biased and unbiased subset of statements. Listeners achieved above-chance accuracy on both subsets of statements (55% accuracy on biased statements, $p = .003$, 56% accuracy on unbiased statements, $p = .038$).

capable of making accurate inferences about speakers' degree of certainty. While Study 2 provides evidence that speakers can effectively use paralinguistic cues to convey their degree of certainty to others when motivated to do so, Study 3 suggests that they are capable of doing so even in the absence of financial motives or constraints to their ability to provide explicit linguistic information about their degree of certainty. In combination, Studies 2 and 3 build on the naturalistic paralanguage examined in Study 1 by finding a causal effect consistent with the notion that perceivers are capable of using speakers' paralanguage to make accurate inferences about their degree of certainty.

5. Meta-Analysis of Vocal Cues

Studies 1-3 have documented that whether through extended naturalistic speech, experimentally induced strategic speech, or a middle ground of the two, speakers encode certainty-signaling cues through paralanguage in a manner that is readily interpreted by listeners. But what specific vocal cues do speakers use to convey certainty? Furthermore, once exposed to speakers' paralanguage, what cues do listeners rely on to infer speakers' certainty? In this section, we answer these questions by presenting the results of a meta-analysis of the cues used by speakers to convey certainty and listeners to infer speakers' certainty.

Prior research on vocal displays of confidence has suggested that speakers' volume, pitch, speech rate, and speech variability may be informative cues to inferring their degree of confidence (Aronovitch 1976, Kimble and Seidel 1991, Scherer et al. 1973). As such, we conducted automated analyses of the audio files obtained from speakers in Studies 1-3 to extract objective measures of their use of these vocal cues. Specifically, we collected data on 9 cues: 1) intensity (volume), 2) intensity slope (|ΔIntensity|), 3) variation in intensity (ΔIntensity), 4) pitch, 5) pitch slope (|ΔPitch|), 6) variation in pitch (ΔPitch), 7) speech rate (Articulation Rate), 8) number of extended pauses (Npauses), and 9) percentage of recording time spent in silence (PctUnvoiced).[5] As the context in which speakers' cues were elicited vary across studies, we focus our discussion of overall effects as opposed to study-specific effects (see Table S2 in SOM for summary statistics of all measures and a correlation matrix).

5.1. The Effect of Speaker Certainty on Speakers' Cues

First, we analyzed the effect of speakers' degree of certainty on their use of cues. To do this, we extracted raw versions of recordings that we obtained from speakers. In Study 1, this represented the non-filtered version of speakers' recordings. In Study 2, this represented the full version of audio we obtained from speakers in the experiment. Because the linguistic content of speakers' recordings in Study 3 differed from Study 2, we analyzed the version of speakers' recordings that omitted the probability phrases

[5] See SOM for a detailed explanation of how measures for each cue were extracted and computed.

preceding statements (i.e., the version presented to listeners). Our goal in doing this was to facilitate a comparison to Study 2. As a further measure in facilitating cross-study comparisons while allowing us to assess relative effect sizes, we standardized speaker certainty and cue measures within each study.

We analyzed the effect of speakers' certainty on cue usage using a three-level model with individual recordings nested within speakers, who were in turn nested within studies. Because our goal is to generalize across speakers and studies, we included speaker-specific and study-specific random intercepts. Furthermore, consistent with the recommendations of Bonett (2009) and Cumming (2013) for analyzing heterogeneous studies, we incorporated a random coefficient that allowed the effect of speaker certainty to vary across studies; this approach relaxes assumptions common to fixed-effects approaches that each study is testing the same effect. Overall, we analyzed 1,577 recordings from 144 different speakers.[6]

As documented in Table 2, we identified a number of effects within studies, but only two effects were reliable across studies. First, speakers' variance in pitch became more pronounced as their degree of certainty increased, $z = 3.01$, $p = .003$. The second effect to emerge was one where participants spoke at a faster rate (i.e., more syllables spoken per unit of time not in silence) to convey certainty, $z = 2.22$, $p = .027$.

Table 2. Meta-Analysis: Effects of Speaker Certainty on Speakers' Vocal Cues

Cue	Study 1	Study 2	Study 3	All Studies
Intensity	-.03 (.06)	.25 (.03)***	.06 (.01)***	.09 (.07)
|ΔIntensity|	-.03 (.06)	.35 (.06)***	.04 (.03)	.12 (.09)
ΔIntensity	.02 (.06)	-.03 (.06)	.06 (.03)*	.04 (.02)
Pitch	-.01 (.05)	.01 (.02)	.01 (.01)	.01 (.01)
|ΔPitch|	.18 (.05)***	.14 (.05)*	.05 (.02)	.10 (.03)**
ΔPitch	.10 (.06)	-.20 (.06)***	-.004 (.03)	-.03 (.07)
Articulation Rate	.09 (.06)	.17 (.04)***	.03 (.03)	.08 (.04)*
Npauses	-.13 (.06)*	.12 (.05)*	.01 (.03)	-.008 (.06)
PctUnvoiced	.02 (.06)	-.22 (.06)***	-.06 (.03)*	-.09 (.05)
N	227	237	1,113	1,577

Note. Numbers represent standardized coefficient estimates of the effect of speakers' certainty on each vocal cue (standard errors in parentheses). The models for Studies 1-3 were estimated using linear regression models with a participant-specific random intercept. The model that combined studies used a three-level structure with individual recordings nested within speakers, who were in turn nested within studies; we included participant-specific and study-specific random intercepts (estimated using maximum likelihood estimation), in addition to a random coefficient that allowed the effect size of speaker certainty to vary across

[6] In Study 3, three speakers' recordings could not be analyzed because their pitch fell below the pitch floor we used for extracting these measures (i.e., their normal speaking pitch was lower than what could be reliably detected with our speech analysis settings). Due to difficulties in analyzing their recordings, we omitted all 36 recordings of these speakers.

studies (estimated using an unstructured covariance matrix).
* $p < .05$. ** $p < .01$. *** $p < .001$.

5.2. The Effect of Speakers' Cues on Listener Perceptions

After analyzing speakers' cues, we conducted a separate set of analyses to assess the impact of particular cues on listeners' inferences about speakers' certainty. To do this, we used the same measures extracted for the analysis of speaker certainty described in Section 5.1 and used them to predict listeners' estimates of speakers' degree of certainty. However, because the version of recordings that we presented to listeners in Study 1 had a constant intensity, we did not include measures of this cue for analyses of Study 1. Furthermore, as we are primarily interested in understanding perceptions of speakers' certainty (as opposed to doubt), we only analyzed the perceptions of listeners in the uncertainty-certainty dimension condition. To facilitate a comparison of effect sizes across studies, we standardized all cue measures and listeners' perceptions of speaker certainty within each study. Overall, we analyzed 2,711 evaluations made by 497 different listeners.

We analyzed the effect of speakers' vocal cues on listeners' perceptions of their degree of certainty using a similar approach to our analyses of the effect of speakers' certainty on their use of vocal cues. Because we were interested in generalizing our findings across individuals and across studies, we used a three-level multilevel linear regression model with recordings nested within listeners, who were in turn nested within studies. To do this, we included study-specific and listener-specific random-intercepts. Because listeners heard recordings of the same speaker in all three studies, the listener-specific random intercept is functionally a random intercept for all listener-speaker dyads. We also incorporated a series of random coefficients that allowed the effect of each vocal cue on listeners' perceptions to vary across studies.

Overall, we found that the same two cues used by speakers to convey certainty predicted listener perceptions (see Table 3). Listeners perceived speakers as conveying a higher degree of certainty as speakers increased their variance in pitch ($z = 3.87$, $p < .001$) and spoke at a faster rate ($z = 4.68$, $p < .001$). As documented in Table S3 of the SOM, the individual effects identified in Studies 2 and 3 hold when using logistic regression models instead.

Table 3. Meta-Analysis: Regressions Predicting Listeners' Perceptions of Speaker Certainty

	Study 1	Study 2	Study 3	Studies 2-3	All Studies
Intensity	----	.06 (.07)	.05 (.03)	.06 (.03)	----
|ΔIntensity|	.03 (.03)	.23 (.07)***	-.05 (.03)	.09 (.11)	.07 (.07)
ΔIntensity	.01 (.03)	.05 (.06)	-.06 (.03)	-.04 (.03)	-.01 (.03)
Mean Pitch	-.03 (.03)	-.03 (.07)	-.04 (.04)	-.04 (.03)	-.03 (.02)

|ΔPitch|	.10 (.04)**	.10 (.07)	.06 (.04)	.07 (.03)*	.09 (.02)***
ΔPitch	.03 (.03)	-.11 (.06)	-.04 (.03)	-.05 (.03)	-.02 (.03)
Articulation Rate	.10 (.03)***	.09 (.07)	.08 (.03)**	.08 (.03)**	.10 (.02)***
Npauses	-.09 (.03)**	.04 (.07)	.01 (.03)	.01 (.03)	-.02 (.04)
PctUnvoiced	.03 (.03)	-.14 (.06)*	-.08 (.03)**	-.10 (.03)***	-.07 (.04)
N	1,361	237	1,113	1,350	2,711
Log Likelihood	-1,899.55	-315.24	-1,566.21	-1,888.60	-3,801.46

Note. Numbers represent standardized coefficient estimates of the effect of speakers' certainty on each vocal cue (standard errors in parentheses). The models for Studies 1-3 were estimated using linear regression models with a listener-specific random intercept. The two models that combine studies use a three-level structure with individual recordings nested within listeners, who were in turn nested within studies; we included listener-specific and study-specific random intercepts (estimated using maximum likelihood estimation), in addition to a random coefficient that allowed the effect size of speaker certainty to vary across studies (estimated using an unstructured covariance matrix).
* $p < .05$. ** $p < .01$. *** $p < .001$.

5.3. Discussion

Though listeners displayed accuracy at detecting speakers' certainty across our three studies, exploratory analyses of speakers' vocal cues revealed that among a variety of cues that we reasonably suspected could be reliable signals of speakers' certainty, only two emerged as reliable. Specifically, as speakers' degree of probabilistic certainty increased, so did their speech rate and variability in their vocal pitch. Critically, listeners were capable of using these cues to guide their judgments of speakers' certainty, which were accurate more often than not.

6. General Discussion

Overall, our data suggests that paralanguage can be a reliable indicator of speakers' degree of probabilistic certainty. Across three studies ranging from extended naturalistic speech to strategic experimentally-manipulated speech, we find that independently of the actual linguistic content of speech, perceivers display accuracy at detecting speakers' certainty in future predictions and in the veracity of information. While scholars in the past have asserted that paralanguage is a critical component of inferring confidence in others (Maslow 1971, Walker 1977), no work to date has directly examined whether people are even capable of using paralanguage to make accurate inferences about lay speakers' confidence. Our work addresses this gap in the literature by providing causal evidence that speakers use paralanguage in a fashion that allows perceivers to accurately infer their degree of probabilistic certainty.

To our knowledge, this represents the first empirical demonstration that people readily use paralanguage to infer a speaker's self-perceived likelihood of accurately forecasting the future or providing factual information. An emerging domain in decision research has been in understanding how vague indicators of probabilistic information like verbal probability statements influence individuals' decisions and perception of others' beliefs relative to more precise numerical statements (e.g., Budescu and Wallsten 1990, Du and Budescu 2005, Piercey 2009, Wallsten and Budescu 1983, Wallsten et al. 1993). The current research documents that paralanguage is yet another vague indicator of probabilistic information—particularly an indicator of speakers' probabilistic certainty. Given the host of social and economic benefits associated with the display of certainty (Hinsz 1990, Radzevick and Moore 2011, Sniezek and Van Swol 2001, Tetlock 2005, Van Swol and Sniezek 2005, Zarnoth and Sniezek 1997), we consider our findings to be of considerable practical value. People use others' displays of certainty to draw inferences about their trustworthiness and credibility (Price and Stone 1994, Phillips 1999, Sah et al. 2013, Tenney et al. 2007, Sniezek and Buckley 1995, Yaniv 1997). While many scholars have documented that they use others' verbal statements of certainty to draw these inferences, our findings raise the possibility that they also use speakers' paralanguage to draw these same inferences. Furthermore, given that people routinely attempt to influence others' perception of their own momentary state for strategic purposes (Andrade and Ho 2009), our finding that speakers are capable of deliberately manipulating their paralanguage to convey certainty (Study 2) suggests that this may be a valuable impression management tactic. By documenting that people use speakers' paralanguage to reliably infer their degree of certainty, the current research may represent a valuable step in understanding how paralanguage may be used in everyday social interaction to influence the distribution of economic and social rewards.

7. References

Anderson, C., S. Brion, D. A. Moore, J. A. Kennedy. 2012. A status-enhancement account of overconfidence. *J.Pers.Soc.Psychol.* **103**(4) 718-735.

Andrade, E. B., T.-H. Ho. 2009. Gaming emotions in social interactions. *J.Consum.Res*. **36**(4) 539-552.

Aronovitch, C. D.. 1976. The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker. *J.Soc.Psychol*. **99**(2) 207-220.

Boersma, P., D. Weenik. 2015. *Praat, a system for doing phonetics by computer*. Version 6.0.08, retrieved from http://www.praat.org.

Bonett, D. G. 2009. Meta-analytic interval estimation for standardized and unstandardized mean differences. *Psychol. Methods.* **14**(3) 225-238.

Brennan, S. E., M. Williams. 1995. The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *J. Mem. Lang.* **34**(3), 383-398.

Budescu, D. V., T. S. Wallsten. 1990. Dyadic decisions with numerical and verbal probabilities. *Organ.Behav.Hum.Dec.* **46**(2) 240-263.

Budescu, D. V., T. S. Wallsten. 1995. *Processing linguistic probabilities: General principles and empirical evidence*. Academic Press.

Cumming, G. 2013. The new statistics: Why and how. *Psychol. Sci.* **25**(1), 7-29.

Du, N., D. V. Budescu. 2005. The effects of imprecise probabilities and outcomes in evaluating investment options. *Manage.Sci.* **51**(12) 1791-1803.

Faul, F., E. Erdfelder, A. Buchner, & A.-G. Lang. 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods.* **41**(4), 1149-1160.

Faul, F., E. Erdfelder, A.-G. Lang, & A. Buchner. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods.* **39**(2), 175-191.

Flynn, F. J., B. M. Staw. 2004. Lend me your wallets: The effect of charismatic leadership on external support for an organization. *Strategic.Manage.J.* **25**(4) 309-330.

Gallo, C. 2012. Barack Obama: A master class in public speaking. *Forbes*. Online edition (November 20).

Gentry, W. A. 2009. Nonverbal Obama: Aside from his words. *BusinessWeek*. Online edition (January 20).

Hinsz, V. B. 1990. Cognitive and consensus processes in group recognition memory performance. *J.Pers.Soc.Psychol.* **59**(4) 705-718.

Human, L. J., J. C. Biesanz. 2013. Targeting the good target: An integrative review of the characteristics and consequences of being accurately perceived. *Pers.Soc.Psychol.Rev.* **17**(3), 248-272.

Kantor, J. 2008. Barack Obama, forever sizing up. *New York Times*. Online edition (October 26).

Kimble, C. E., S. D. Seidel. 1991. Vocal signs of confidence. *J.Nonverbal.Behav.* **15**(2) 99-105.

Maslow, C., K. Yoselson, H. London. 1971. Persuasiveness of confidence expressed via language and body language.*Brit.J.Soc.Clin.Psyc.* **10**(3) 234-240.

Meindl, J. R., S. B. Ehrlich, J. M. Dukerich. 1985. The romance of leadership. *Admin.Sci.Quart.* **30**(1) 78-102.

Moore, D. A., P. J. Healy. 2008. The trouble with overconfidence. *Psychol.Rev.* **115**(2) 502-517.

Piercey, M. D. 2009. Motivated reasoning and verbal vs. numerical probability assessment: Evidence from an accounting context. *Organ. Behav. Hum. Dec.* **108**(2), 330-341.

Price, P. C., E. R. Stone. 2004. Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *J.Behav.Decis.Making*. **17**(1) 39-57.

Phillips, J. M. 1999. Antecedents of leader utilization of staff input in decision-making teams. *Organ. Behav. Hum. Dec.* **77**(3), 215-242.

Radzevick, J. R., D. A. Moore. 2011. Competing to be certain (but wrong): Market dynamics and excessive confidence in judgment. *Manage.Sci.* **57**(1) 93-106.

Ramirez Verdugo, D.. 2005. The nature and patterning of native and non-native intonation in the expression of certainty and uncertainty: Pragmatic effects. *J.Pragmatics.* **37**(12) 2086-2115.

Sah, S., D. A. Moore, & R. J. MacCoun. 2013. Cheap talk and credibility: The consequences of confidence and accuracy on advisor credibility and persuasiveness. *Organ. Behav. Hum. Dec.* **121**(2), 246-255.

Scherer, K. R. 1971. Randomized splicing: A note on a simple technique for masking speech content. *J. Exp. Res. Pers.* **5**(2), 155-159.

Scherer, K. R., D. R. Ladd, K. E. A. Silverman. 1984. Vocal cues to speaker affect: Testing two models. *J. Acoust. Soc. Am.* **76**(5), 1346-1356.

Scherer, K. R., H. London, J. J. Wolf. 1973. The voice of confidence: Paralinguistic cues and audience evaluation. *J.Res.Pers.* **7**(1) 31-44.

Smith, V. L., & H. H. Clark. 1993. On the course of answering questions. *J. Mem. Lang.* **32**(1), 25-38.

Sniezek, J. A., & T. Buckley. 1995. Cueing and cognitive conflict in judge-advisor decision making. *Organ. Behav. Hum. Dec.* **62**(2), 159-174.

Sniezek, J. A., L. M. Van Swol. 2001. Trust, confidence, and expertise in a judge-advisor system. *Organ.Behav.Hum.Dec.* **84**(2) 288-307.

Tenney, E. R., R. J. MacCoun, B. A. Spellman, R. Hastie. 2007. Calibration trumps confidence as a basis for witness credibility. *Psychol.Sci.* **18**(1) 46-50.

Tenney, E. R., B. A. Spellman, R. J. MacCoun. 2008. The benefits of knowing what you know (and what you don't): How calibration affects credibility. *J. Exp. Soc. Psychol.* **44**(5), 1368-1375.

Tetlock, P. E. 2005. *Expert political judgment: How good is it? How can we know?* Princeton University Press, Princeton, NJ.

Traynham, R. L. 2012. Racism drives some Obama critics. *Philadelphia Tribune*. Online edition (June 24).

Valley, K. L., S. B. White, M. A. Neale, M. H. Bazerman. 1992. Agents as information brokers: The effects of information disclosure on negotiated outcomes. *Organ. Behav. Hum. Dec.* **51**(2), 220-236.

Van Swol, L. M., J. A. Sniezek. 2005. Factors affecting the acceptance of expert advice. *Brit.J.Soc.Psychol.* **44**(3) 443-461.

Van Zant, A. B., D. A. Moore. 2013. Avoiding the pitfalls of overconfidence while benefitting from the advantages of confidence. *Calif. Manage. Rev.* **55**(2), 5-23.

Walker, M. B. 1977. The relative importance of verbal and nonverbal cues in the expression of confidence. *Aust.J.Psychol.* **29**(1) 45-57.

Wallsten, T. S., D. V. Budescu. 1983. State of the art—Encoding subjective probabilities: A psychological and psychometric review. *Manage. Sci.* **29**(2), 151-173.

Westley, F., H. Mintzberg. 1989. Visionary leadership and strategic management. *Strateg.Manage.J.* **10**(S1) 17-32.

Zarnoth, P., J. A. Sniezek. 1997. The social influence of confidence in group decision making. *J.Exp.Soc.Psychol.* **33**(4) 345-366.

Appendix A

Statements Used for Studies 1 and 2

Statement Set A	Statement Set B
Statement 1	*Statement 1*
James Monroe was 73 years old when he died. (*True*)	You can't sink in quicksand if you raise your legs slowly and lie on your back. (*True*)
James Monroe was 75 years old when he died. (*False*)	You can't sink in quicksand if you raise your arms slowly and lie on your back. (*False*)
Statement 2	*Statement 2*
There are currently 23 whiskey distilleries in Germany. (*True*)	Oregon has more ghost towns than any other state. (*True*)
There are currently 27 whisky distilleries in Germany. (*False*)	Massachusetts has more ghost towns than any other state. (*False*)
Statement 3	*Statement 3*
The modern felt-tip pen was invented in Japan. (*True*)	The average person walks the equivalent of five times around the equator in a lifetime. (*True*)
The modern felt-tip pen was invented in China. (*False*)	The average person walks the equivalent of three times around the equator in a lifetime. (*False*)
Statement 4	*Statement 4*
The waterproof wristwatch was invented by Rolex in 1923. (*True*)	The most common name in the world is Mohammed. (*True*)
The waterproof wristwatch was invented by Rolex in 1933. (*False*)	The most common name in the world is John. (*False*)
Statement 5	*Statement 5*
Two people from Connecticut signed the Declaration of Independence. (*True*)	The most money a cow has ever sold for in an auction is $1.3 million. (*True*)
One person from Connecticut signed the Declaration of Independence. (*False*)	The most money a cow has ever sold for in an auction is $100,000. (*False*)
Statement 6	*Statement 6*
No piece of paper can be folded in half more than 7 times. (*True*)	The Mona Lisa is 30 inches by 21 inches. (*True*)
No piece of paper can be folded in half more than 12 times. (*False*)	The Mona Lisa is 37 inches by 30 inches. (*False*)

Appendix B

Instructions for Participants' Selection of Probability Phrases in Study 2

People often use specific phrases to indicate that do not know whether something is true or false. Suppose you know that the statement "Joe was born in California" [***is*** / ***may be***] ***true***. That is, you know there to be a [100% / 50%] probability that Joe was born in California.

If you had to relay this information to somebody else, which of the following phrases would you use to describe what you know about the likelihood that Joe was born in California? Please select one of the phrases below and click "Continue".

Choices for 100% probability screen:

- I am certain that Joe was born in California.
- I am sure that Joe was born in California.
- I am positive that Joe was born in California.
- I am confident that Joe was born in California.
- I am convinced that Joe was born in California.
- I can assure you that Joe was born in California.
- I can guarantee that Joe was born in California.
- It is clear that Joe was born in California.

Choices for 50% probability screen:

- I am uncertain whether Joe was born in California.
- I am unsure whether Joe was born in California.
- I have no idea whether Joe was born in California.
- I cannot determine whether Joe was born in California.
- It is a toss-up whether Joe was born in California.
- It is questionable whether Joe was born in California.
- It is unclear whether Joe was born in California.
- It is debatable whether Joe was born in California.