# In Search of a Gold Standard in Studies of Deception

**Stephanie Gokhman**[1], **Jeff Hancock**[1,3], **Poornima Prabhu**[2], **Myle Ott**[2], **Claire Cardie**[2,3]
Departments of Communication[1], Computer Science[2], and Information Science[3]
Cornell University, Ithaca, NY 14853
{sbg94,jth34,pmp67,mao37,ctc9}@cornell.edu

**Abstract**

In this study, we explore several popular techniques for obtaining corpora for deception research. Through a survey of traditional as well as non-gold standard creation approaches, we identify advantages and limitations of these techniques for web-based deception detection and offer crowdsourcing as a novel avenue toward achieving a gold standard corpus. Through an in-depth case study of online hotel reviews, we demonstrate the implementation of this crowdsourcing technique and illustrate its applicability to a broad array of online reviews.

## 1 Introduction

Leading deception researchers have recently argued that verbal cues are the most promising indicators for detecting deception (Vrij, 2008) while lamenting the fact that the majority of previous research has focused on nonverbal cues. At the same time, increasing amounts of language are being digitized and stored on computers and the Internet — from email, Twitter and online dating profiles to legal testimony and corporate communication. With the recent advances in natural language processing that have enhanced our ability to analyze language, researchers now have an opportunity to similarly advance our understanding of deception.

One of the crucial components of this enterprise, as recognized by the call for papers for the present workshop, is the need to develop corpora for developing and testing models of deception. To date there has not been any systematic approach for corpus creation within the deception field. In the present study, we first provide an overview of traditional approaches for this task (Section 2) and discuss recent deception detection methods that rely on non-gold standard corpora (Section 3). Section 4 introduces novel approaches for corpus creation that employ crowdsourcing and argues that these have several advantages over traditional and non-gold standard approaches. Finally, we describe an in-depth case study of how these techniques can be implemented to study deceptive online hotel reviews (Section 5).

## 2 Traditional Approaches

The deception literature involves a number of widely used traditional methods for gathering deceptive and truthful statements. We classify these according to whether they are *sanctioned*, in which the experimenter supplies instructions to individuals to lie or not lie, or *unsanctioned* approaches, in which the participant lies of his or her own accord.

### 2.1 Sanctioned Deception

The vast majority of studies examining deception employ some form of the sanctioned lie method. A common example is recruiting participants for a study on deception and randomly assigning them to a lie or truth condition. A classic example of this kind of procedure is the original study by Ekman and Friesen (1969), in which nurses were required to watch pleasant or highly disturbing movie clips. The nurses were instructed to indicate that they were watching a pleasing movie, which required the nurses watching the disturbing clips to lie about their current emotional state.

In another example, Newman et. al. (2003) ask



*Proceedings of the EACL 2012 Workshop on Computational Approaches to Deception Detection*, pages 23–30,
Avignon, France, April 23 - 27 2012. 

participants about their beliefs concerning a given topic, such as abortion, and then instruct participants to convince a partner that they hold the opposite belief.

Another form of sanctioned deception is to instruct participants to engage in some form of mock crime and then ask them to lie about it. For example, in one study (Porter and Yuille, 1996), participants were asked to take an item, such as a wallet, from a room and then lie about it afterwards. The mock crime approach improves the ecological validity of the deception, and makes it the case that the person actually did in fact act a certain way that they then must deny.

#### 2.1.1 Advantages and Limitations

The advantages are obvious for these sanctioned lie approaches. The researcher has large degrees of experimental control over what the participant lies about and when, which allows for careful comparison across the deceptive and non-deceptive accounts. Another advantage is the relative ease of instructing participants to lie vs. trying to identify actual (but unknown) lies in a dialogue.

The limitations for this approach, however, are also obvious. In asking participants to lie, the researcher is essentially giving permission to the person to lie. This should affect the participant's behavior as the lie is being conducted at the behest of a power figure, essentially acting out their deception. Indeed, a number of scholars have pointed out this problem (Frank and Ekman, 1997), and have suggested that unless high stakes are employed the paradigm produces data that does not replicate any typical lying situation. *High stakes* refers to the potential for punishment if the lie is detected or reward if the lie goes undetected. Perhaps because of the difficulty in creating high-stakes deception scenarios, to date there are few corpora involving high-stakes lies.

### 2.2 Unsanctioned Deception

Unsanctioned lies are those that are told without any explicit instruction or permission from the researcher. These kinds of lies have been collected in a number of ways.

#### 2.2.1 Diary studies and surveys

Two related methods for collecting information about unsanctioned lies are diary studies and survey studies. In diary studies participants are asked on an ongoing basis (e.g., every night) to recall lies that they told over a given period (e.g., a day, a week) (DePaulo et al., 1996; Hancock et al., 2004). Similarly, recent studies have asked participants in national surveys how often they have lied in the last 24 hours (Serota et al., 2010).

One important feature of these approaches is that the lies have already taken place, and thus they do not share the same limitations as sanctioned lies. There are several drawbacks, however, especially given the current goal to collect deception corpora. First, both diary studies and survey approaches require self-reported recall of deception. Several biases are likely to affect the results, including under-reporting of deception in order to reduce embarrassment and difficult-to-remember deceptions that have occurred over the time period. More importantly, this kind of approach does not lend itself to collecting the actual language of the lie, for incorporation into a corpus: people have a poor memory for conversation recall (Stafford and Sharkey, 1987).

#### 2.2.2 Retrospective Identification

One method for getting around the memory limitations for natural discourse is to record the discourse and ask participants to later identify any deceptions in their discourse. For instance, one study (Feldman and Happ, 2002) asked participants to meet another individual and talk for ten minutes. After the discussion, participants were asked to examine the videotape of the discussion and indicated any times in which they were deceptive. More recently, others have used the retrospective identification technique on mediated communication, such as SMS, which produces an automatic record of the conversation that can be reviewed for deception (Hancock, 2009). Because this approach preserves a record that the participant can use to identify the deception, this technique can generate data for linguistic analysis. However, an important limitation, as with the diary and survey data, is that the researcher must assume that the participant is being truthful about their deception reporting.

#### 2.2.3 Cheating Procedures

The last form of unsanctioned lying involves incentivizing participants to first cheat on a task and to then lie when asked about the cheating behavior. Levine et al. (2010) have recently used this approach, which involved students performing a trivia quiz. During the quiz, an opportunity to cheat arises where some of the students will take the opportunity. At this point, they have not yet lied, but, after the quiz is over, all students are asked whether they cheated by an interviewer who does not know if they cheated or not. While most of the cheaters admit to cheating, a small fraction of the cheaters deny cheating. This subset of cheating denials represents real deception.

The advantages to this approach are threefold: (1) the deception is unsanctioned, (2) it does not involve self-report, and (3) the deceptions have objective ground-truth. Unfortunately, these kinds of experiments are extremely effort-intensive given the number of deceptions produced. Only a tiny fraction of the participants typically end up cheating and subsequently lying about the cheating.

#### 2.2.4 Limitations

While these techniques have been useful in many psychology experiments, in which assessing deception detection has been the priority rather than corpus creation, they are not very feasible when considering obtaining corpora for large-scale settings, e.g., the web. Furthermore, the techniques are limited in the kinds of contexts that can be created. For instance, in many cases, e.g., deliberate posting of fake online reviews, subjects can be both highly incentivized to lie and highly concerned with getting caught. One could imagine surveying hotel owners as to whether they have ever posted a fake review—but it would seem unlikely that any owner would ever admit to having done so.

## 3 Non-gold Standard Approaches

Recently, alternative approaches have emerged to study deception in the absence of gold standard deceptive data. These approaches can typically be broken up into three distinct types. In Section 3.1, we discuss approaches to deception corpus creation that rely on the *manual annotation* of deceptive instances in the data. In Section 3.2, we discuss approaches that rely on *heuristic methods* for deriving approximate, but non-gold standard deception labels. In Section 3.3, we discuss a recent approach that uses assumptions about the effects of deception to identify examples of deception in the data. We will refer to the latter as the *unlabeled* approach to deception corpus creation.

### 3.1 Manual Annotations of Deception

In Section 2.2, we discussed diary and self-report methods of obtaining gold standard labels of deception. Recently, work studying deceptive (fake) online reviews has suggested using manual annotations of deception, given by third-party human judges.

Lim et al. (2010) study deceptive product reviews found on Amazon.com. They develop a sophisticated software interface for manually labeling reviews as deceptive or truthful. The interface allows annotators to view all of each user's reviews, ranked according to dimensions potentially of importance to identifying deception, e.g., whether the review is duplicated, whether the reviewer has authored many reviews in a single day with identical high or low ratings, etc.

Wu et al. (2010a) also study deceptive online reviews of TripAdvisor hotels, manually labeling a set of reviews according to "suspiciousness." This manually labeled dataset is then used to validate eight proposed characteristics of deceptive hotels. The proposed characteristics include features based on the number of reviews written, e.g., by first-time reviewers, as well as the review ratings, especially as they compare to other ratings of the same hotel.

Li et al. (2011) study deceptive product reviews found on Epinions.com. Based on user-provided helpfulness ratings, they first draw a subsample of reviews such that the majority are considered to be unhelpful. They then manually label this subsample according to whether or not each review seems to be fake.

#### 3.1.1 Limitations

Manual annotation of deception is problematic for a number of reasons. First, many of the same challenges that face manual annotation efforts in other domains also applies to annotations of deception. For example, manual annotations can be expensive to obtain, especially in large-scale settings, e.g., the web.

Most seriously however, is that human ability to detect deception is notoriously poor (Bond and DePaulo, 2006). Indeed, recent studies have confirmed that human agreement and deception detection performance is often no better than chance (Ott et al., 2011); this is especially the case when considering the overtrusting nature of most human judges, a phenomenon referred to in the psychological deception literature as a truth bias (Vrij, 2008).

### 3.2 Heuristically Labeled

Work by Jindal and Liu (2008) studying the characteristics of untruthful (deceptive) Amazon.com reviews, has instead developed an approach for *heuristically* assigning approximate labels of deceptiveness, based on a set of assumptions specific to their domain. In particular, after removing certain types of irrelevant "reviews," e.g., questions, advertisements, etc., they determine whether each review has been duplicated, i.e., whether the review's text heavily overlaps with the text of other reviews in the same corpus. Then, they simply label all discovered duplicate reviews as untruthful.

Heuristic labeling approaches do not produce a true gold-standard corpus, but for some domains may offer an acceptable approximation. However, as with other non-gold standard approaches, certain behaviors might have other causes, e.g., duplication could be accidental, and just because something is duplicated does not make the original (first) post deceptive. Indeed, in cases where the original review is truthful, its duplication is not a good example of deceptive reviews written from scratch.

### 3.3 Unlabeled

Rather than develop heuristic labeling approaches, Wu et al. (2010b) propose a novel strategy for evaluating hypotheses about deceptive hotel reviews found on TripAdvisor.com, based on distortions of popularity rankings. Specifically, they test the *Proportion of Positive Singletons* and *Concentration of Positive Singletons* hypotheses of Wu et al. (2010a) (Section 3.1), but instead of using manually-derived labels they evaluate their hypotheses by the corresponding (distortion) effect they have on the hotel rankings.

Unlabeled approaches rely on assumptions about the effects of the deception. For example, the approach utilized by Wu et al. (2010b) observing distortion effects on hotel rankings, relies on the assumption that the goal of deceivers in the online hotel review setting is to increase a hotel's ranking. And while this may be true for positive hotel reviews, it is likely to be very *untrue* for fake negative reviews intended to defame a competitor. Indeed, great care must be taken in making such assumptions in unlabeled approaches to studies of deception.

## 4 Crowdsourcing Approaches

As with traditional sanctioned deception approaches (see Section 2.1), one way of obtaining gold standard labels is to simply create gold standard deceptive content. Crowdsourcing platforms are a particularly compelling space to produce such deceptive content: they connect people who request the completion of small tasks with workers who will carry out the tasks. Crowdsourcing platforms that solicit small copywriting tasks include Clickworker, Amazon's Mechanical Turk, Fiverr, and Worth1000. Craigslist, while not a crowdsourcing platform, also promotes similar solicitations for writing. In the case of fake online reviews (see Section 5), and by leveraging platforms such as Mechanical Turk, we can often generate gold standard deceptive content in contexts very similar to those observed in practice.

Mihalcea and Strapparava (2009) were among the first to use Mechanical Turk to collect deceptive and truthful opinions — personal stances on issues such as abortion and the death penalty. In particular, for a given topic, they solicited one truthful and one deceptive stance from each Mechanical Turk participant.

Ott et al. (2011) have also used Mechanical Turk to produce gold standard deceptive content. In particular, they use Mechanical Turk to generate a dataset of 400 *positive* (5-star), gold standard deceptive hotel reviews. These were combined with 400 (positive) *truthful* reviews covering the same set of hotels and used to train a learning-based classifier that could distinguish deceptive vs. truthful positive reviews at 90% accuracy levels. The truthful reviews were mined directly from a well-known hotel review site. The Ott et al. (2011) approach for collecting the gold standard deceptive reviews is the subject of the case study below.

## 5 Case Study: Crowdsourcing Deceptive Reviews

To illustrate in more detail how crowdsourcing techniques can be implemented to create gold standard data sets for the study of deception, we draw from the Ott et al. (2011) approach that crowdsources the collection of **deceptive positive hotel reviews** using Mechanical Turk. The key assumptions of the approach are as follows:

- We desire a **balanced data set**, i.e., equal numbers of truthful and deceptive reviews. This is so that statistical analyses of the data set won't be biased towards either type of review.

- The truthful and deceptive reviews should **cover the same set of entities**. If the two sets of reviews cover different entities (e.g., different hotels), then the language that distinguishes truthful from deceptive reviews might be attributed to the differing entities under discussion rather than to the legitimacy of the review.

- The resulting data set should be of a **reasonable size**. Ott et al. (2011) found that a dataset of 800 total reviews (400 truthful, 400 deceptive) was adequate for their goal of training a learning-based classifier.

- The truthful and deceptive reviews should **exhibit the same valence, i.e., sentiment**. If the truthful reviews gathered from the online site are *positive* reviews, the deceptive reviews should be positive as well.

- More generally, the **deceptive reviews should be generated under the same basic guidelines as governs the generation of truthful reviews**. E.g., they should have the same length constraints, the same quality constraints, etc.

**Step 1: Identify the set of entities to be covered in the** ***truthful*** **reviews.** In order to define a set of desirable reviews, a master database, provided by the review site itself, is mined to identify the most commented (most popular) entities. These are a good source of *truthful* reviews. In particular, previous work has hypothesized that popular offerings are less likely to be targeted by spam (Jindal and Liu, 2008), and therefore reviews for those entities are less likely to be deceptive—enabling those reviews to later comprise the truthful review corpus. The review site database typically divides the entity set into subcategories that differ across contexts: in the case of hotel reviews the subcategories might refer to cities, or in the case of doctor reviews subcategories might refer to specialties. To ensure that enough reviews of the entity can be collected, it may be important to select subcategories that themselves are popular. The study of Ott et al. (2011), for example, focused on reviews of hotels in Chicago, IL, gathering positive (i.e., 5-star) reviews for the 20 most popular hotels.

**Step 2: Develop the crowdsourcing prompt.** Once a set of entities has been identified for the deceptive reviews (Step 1), the prompt for Mechanical Turk is developed. This begins with a survey of other solicitations for reviews within the same subcategory through searching Mechanical Turk, Craigslist, and other online resources. Using those solicitations as reference, a scenario can then be developed that will be used in the prompt to achieve the appropriate (in our case, *positive*) valence. The result is a prompt that mimics the vocabulary and tone that "Turkers" (i.e., the workers on Mechanical Turk) may find familiar and desirable.

For example, the prompt of Ott et al. (2011) read: *Imagine you work for the marketing department of a hotel. Your boss asks you to write a fake review for the hotel (as if you were a customer) to be posted on a travel review website.* ***The review needs to sound realistic and portray the hotel in a positive light.*** *Look at their website if you are not familiar with the hotel.* (A link to the website was provided.)

**Step 3: Attach appropriate warnings to the crowdsource solicitation.** It is important that warnings are attached to the solicitation to avoid gathering (and paying for) reviews that would invalidate the review set for the research. For example, because each review should be written by a different person, the warning might disallow coders from performing multiple reviews; forbid any form of plagiarism; require that reviews be "on topic," coherent, etc. Finally, the prompt may inform the Turker that this exercise is for academic purposes only and will not be posted online, however, if such a notice is presented before the review is written and submitted, the resulting lie may be overly sanctioned.

**Step 4: Incorporate into the solicitation a means for gathering additional data.** Append to the end of the solicitation some mechanism (e.g., Mechanical Turk allows for a series of radio buttons) to input basic information about age, gender, or education of the coder. This allows for post-hoc understanding of the demographic of the participating Turkers. Ott et al. (2011) also supply a space for comments by the workers, with an added incentive of a potential bonus for particularly helpful comments. Ott et al. (2011) found this last step critical to the iterative process for providing insights from coders on inconsistencies, technical difficulties, and other unforeseen problems that arise in the piloting phase.

**Step 5: Gather the deceptive reviews in batches.** The solicitation is then published in a small pilot test batch. In Ott et al. (2011), each pilot requested ten (10) reviews from unique workers. Once the pilot run is complete, the results are evaluated, with particular attention to the comments, and is then iterated upon in small batches of 10 until there are no technical complaints and the results are of desired experiment quality.

Once this quality is achieved, the solicitation is then published as a full run, generating 400 reviews by unique workers. The results are manually evaluated and cleaned to ensure all reviews are valid, then filtered for plagiarism. The resulting set of gold standard online deceptive spam is then used to train the algorithm for deceptive positive reviews.

### 5.1 Handling Plagiarism

One of the main challenges facing crowdsourced deceptive content is identifying plagiarism. For example, when a worker on Mechanical Turk is asked to write a deceptive hotel review, that worker may copy an available review from various sources on the Internet (e.g., TripAdvisor). These plagiarized reviews lead to flaws in our gold standard. Hence there arises a need to detect such reviews and separate them from the entire review set.

One way to address this challenge is to do a manual check of the reviews, one-by-one, using online plagiarism detection web services, e.g., plagiarisma.net or searchenginereports.net. The manual process is taxing, especially when there are reviews in large numbers (as large as 400) to be processed. This illustrates a need to have a tool which automates the detection of plagiarized content in Turker submissions. There are several plagiarism detection softwares which are widely available in the market. Most of them maintain a database of content against which to check for plagiarism. The input content is checked against these databases and the content is stored in the same database at the end of the process. Such tools are an appropriate fit for detecting plagiarized content in term papers, course assignments, journals etc. However, online reviews define a separate need which checks for plagiarism against the content available on the web. Hence the available software offerings are not adequate.

We implemented a command line tool using the Yahoo! BOSS API, which is used to query sentences on the web. Each of the review files is parsed to read as individual sentences. Each sentence is passed as a query input to the API. We introduce the parameters, $n$ and $m$, defined as:

1. Any sentence which is greater than $n$ words is considered to be a "long sentence" in the application usage. If the sentence is a "long sentence" and the Yahoo! BOSS API returns no result, we query again using the first $n$ words of the sentence. Here $n$ is a configurable parameter, and in our experiments we configured $n = 10$.

2. A sentence that is commonly used on the web can return many matches, even if it was not plagiarized. Thus, we introduce another parameter, $m$, such that if the number of search results returned by the Yahoo! BOSS API is greater than $m$, then the sentence is considered common and is ignored. Our observations indicate that such frequently used sentences are likely to be short. For example: "We are tired," "No room," etc. For our usage we configured $m = 30$.

We consider a sentence to be plagiarized if the total number of results returned by the Yahoo! BOSS API is less than $m$. Hence each sentence is assigned a score as follows:

- *If the total number of results is greater than $m$: assign a score of 0*

- *If the total number of results is less than or equal to $m$: assign a score of 1*

We then divide the sum of the sentence scores in a review by the total number of sentences to obtain the ratio of the number of matches to total number of sentences. We use this ratio to determine whether or not a review was plagiarized.

## 6 Discussion and Conclusion

We have discussed several techniques for creating and labeling deceptive content, including traditional, non-gold standard, and crowdsourced approaches. We have also given an illustrative in-depth look at how one might use crowdsourcing services such as Mechanical Turk to solicit deceptive hotel reviews.

While we argue that the crowdsourcing approach to creating deceptive statements has tremendous potential, there remain a number of important limitations, some shared by the previous traditional methods laid out above. First, workers are given “permission” to lie, so these lies are sanctioned and have the same concerns as the traditional sanctioned methods, including the concern that the workers are just play-acting rather than lying. Other unique limitations include the current state of knowledge about workers. In a laboratory setting we can fairly tightly measure and control for gender, race, and even socioeconomic status, but this is not the case for the Amazon Turkers, who potentially make up a much more diverse population.

Despite these issues we believe that the approach has much to offer. First, and perhaps most importantly, the deceptions are being solicited in exactly the manner real-world deceptions are initiated. This is important in that the deception task, though sanctioned, is precisely the same task that a real-world deceiver might use, e.g., to collect fake hotel reviews for themselves. Second, this approach is extremely cost effective in terms of the time and finances required to create custom deception settings that fit a specific context. Here we looked at creating fake hotel reviews, but we can easily apply this approach to other types of reviews, including reviews of medical professionals, restaurants, and products.

## Acknowledgments

This work was supported in part by National Science Foundation Grant NSCC-0904913, and the Jack Kent Cooke Foundation. We also thank the EACL reviewers for their insightful comments, suggestions and advice on various aspects of this work.

## References

C.F. Bond and B.M. DePaulo. 2006. Accuracy of deception judgments. *Personality and Social Psychology Review*, 10(3):214.

B.M. DePaulo, D.A. Kashy, S.E. Kirkendol, M.M. Wyer, and J.A. Epstein. 1996. Lying in everyday life. *Journal of personality and social psychology*, 70(5):979.

P. Ekman and W. V. Friesen. 1969. *Nonverbal Leakage And Clues To Deception*, volume 32.

Forrest J. A. Feldman, R. S. and B. R. Happ. 2002. Self-presentation and verbal deception: Do self-presenters lie more? *Basic and Applied Social Psychology*, 24:163–170.

M.G. Frank and P. Ekman. 1997. The Ability To Detect Deceit Generalizes Across Different Types of High-Stake Lies. *Journal of Personality and Social Psychology*, 72:1429–1439.

J.T. Hancock, J. Thom-Santelli, and T. Ritchie. 2004. Deception and design: The impact of communication technology on lying behavior. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 129–134. ACM.

J.T. Hancock. 2009. Digital Deception: The Practice of Lying in the Digital Age. *Deception: Methods, Contexts and Consequences*, pages 109–120.

N. Jindal and B. Liu. 2008. Opinion spam and analysis. In *Proceedings of the international conference on Web search and web data mining*, pages 219–230. ACM.

Kim R. K. Levine, T. R. and J. P. Blair. 2010. (In)accuracy at detecting true and false confessions and denials: An initial test of a projected motive model of veracity judgments. *Human Communication Research*, 36:81–101.

F. Li, M. Huang, Y. Yang, and X. Zhu. 2011. Learning to identify review spam. In *Twenty-Second International Joint Conference on Artificial Intelligence*.

E.P. Lim, V.A. Nguyen, N. Jindal, B. Liu, and H.W. Lauw. 2010. Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 939–948. ACM.

R. Mihalcea and C. Strapparava. 2009. The lie detector: Explorations in the automatic recognition of deceptive language. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 309–312. Association for Computational Linguistics.

M.L. Newman, J.W. Pennebaker, D.S. Berry, and J.M. Richards. 2003. Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29(5):665.

M. Ott, Y. Choi, C. Cardie, and J.T. Hancock. 2011. Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 309–319. Association for Computational Linguistics.

S. Porter and J.C. Yuille. 1996. The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior*, 20:443–458.

K.B. Serota, T.R. Levine, and F.J. Boster. 2010. The prevalence of lying in america: Three studies of self-reported lies. *Human Communication Research*, 36(1):2–25.

Burggraf C. S. Stafford, L. and W.F. Sharkey. 1987. Conversational Memory The Effects of Time, Recall, Mode, and Memory Expectancies on Remembrances of Natural Conversations. *Human Communication Research*, 14:203–229.

A. Vrij. 2008. *Detecting lies and deceit: Pitfalls and opportunities*. Wiley-Interscience.

G. Wu, D. Greene, B. Smyth, and P. Cunningham. 2010a. Distortion as a validation criterion in the identification of suspicious reviews. In *Proceedings of the First Workshop on Social Media Analytics*, pages 10–13. ACM.

G. Wu, D. Greene, B. Smyth, and P. Cunningham. 2010b. Distortion as a validation criterion in the identification of suspicious reviews. Technical report, UCD-CSI-2010-04, University College Dublin.