



AFRL-OSR-VA-TR-2015-0024

IDENTIFYING UNIVERSAL LINGUISTIC FEATURES ASSOCIATED WITH VERACITY AND DECEPTION

David Matsumoto
HUMINTELL LLC

01/02/2015
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTC
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY)		2. REPORT TYPE		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (include area code)

24 December 2014

Dr. Benjamin Knott
Benjamin A Knott, PhD
Air Force Office of Scientific Research
Program Officer - Trust and Influence
875 N. Randolph St.
Arlington, VA 22203
Phone: 703-696-1142
Email: benjamin.knott.2@us.af.mil
Reference: AFOSR FA9550-11-1-0306

Dear Dr. Knott;

In accordance with AFOSR Award No. FA9550-11-1-0306 from AFOSR to Humintell, LLC, and the Contract Data Requirements List (CDRL) of the referenced contract, Humintell is pleased to submit the Final Report for the project. If you have any questions concerning the contents of this report, please contact me at (510) 704-1883.

Respectfully,

A handwritten signature in black ink, reading "David Matsumoto". The signature is fluid and cursive, with a stylized "D" and "M".

Humintell, LLC
David Matsumoto
President

Table of Contents

Project Overview	3
Summary of Specific Findings across the Entire Grant Period	4
Year 1	4
Year 2	5
Year 3	5
Potential Impact or Translation to Military Applications	7
Performance Metrics	8
Bibliography	9
Scientific Publications	9
Transitional Publications	9
Conference Presentations	9
Appendices	10
Appendix 1 – Report on Pilot Study 1, Year 1 (in press at the <i>International Journal of Psychology</i>)	11
Appendix 2 - Report on Main Study, Year 2 (in press at <i>Journal of Police and Criminal Psychology</i>)	31
Appendix 3 - Report on Main Study, Year 3 (currently submitted to a peer-reviewed journal)	58

Project Overview

The ability to exert influence on individuals and groups depends on the ability to make accurate judgments of the veracity of what one is told. Such judgments are at the heart of any interpersonal or intercultural interaction, and contribute to the development or rapport; guides the nature of influence, negotiation, vetting, and information collection; and the development of trust. Because of their importance, there is an abundant literature on the cues to deception, and based on this literature, there have been a number of techniques developed over the years to evaluate truth and detect deception. One important genre of such techniques involves the analysis of verbal statements, collectively known as Statement Analysis (SA).

Different types of SA techniques exist, and research has demonstrated that *all* of them are able to detect truths from lies at better than chance accuracies, and in different languages. This suggests that there may be something *universal* to SA. This notion receives support from knowledge concerning the universal principles and processes of memory encoding, as well as the deep structure of language.

But to date there has never been a study of SA across multiple languages using a standard paradigm that examines its cross-cultural applicability. This three-year project addressed this gap in the literature. Year 1 involved the conduct of two pilot studies to ensure the validity of the stimuli and procedures used in the main study, conducted in Year 2. The Year 2 study involved an eyewitness testing paradigm, in which participants from three very different language groups witnessed an actual crime and write true and false witness statements about what they saw. Year 3 involved a replication of the findings from Year 2 in a completely different research paradigm – a mock crime. Multilingual SA experts with years of field experience coded the statements from both Years 2 and 3. We hypothesized that false statements will include significantly more indicators of deception and significantly less indicators of veracity across all languages tested.

Summary of Specific Findings across the Entire Grant Period

Year 1

Pilot Study 1. The purpose of this study was to ensure the validity and cross-cultural equivalence of the crime videos that were to be used in Year 2. Observers in seven countries viewed seven videos portraying actual crimes and rated their emotional reactions to each using 14 emotion scales. Observers reported significantly high levels of negative emotions including anger, contempt, disgust, fear, and sadness-related emotions, and anger, contempt, and disgust were the most salient emotions experienced by viewers across all countries. Witnesses also reported significantly high levels of positive emotions

Table 1

ICCs across the 14 Emotion Means using Countries as Raters

Video	ICC for absolute agreement	ICC for consistency
1	0.890	0.930
2	0.902	0.937
3	0.910	0.941
4	0.913	0.948
5	0.906	0.945
6	0.910	0.944
7	0.927	0.957
All	0.913	0.946

as well (compared to not feeling the emotion at all), which was unexpected. There was considerably high agreement across the countries on the relative rankings across means of the 14 emotions rated for each of the seven videos and across all videos overall (Table 1). These findings suggested that overall there was a great deal of consistency across the countries in the means of their emotional profiles for each of the videos. We then selected the video(s) with the

highest emotional impact (highest overall emotion ratings) to use in Year 2.

Pilot Study 2. The purpose of Pilot Study 2 was to examine possible differential carry-over effects if participants wrote both true and false statements in a within-subjects design. The findings were as follows:

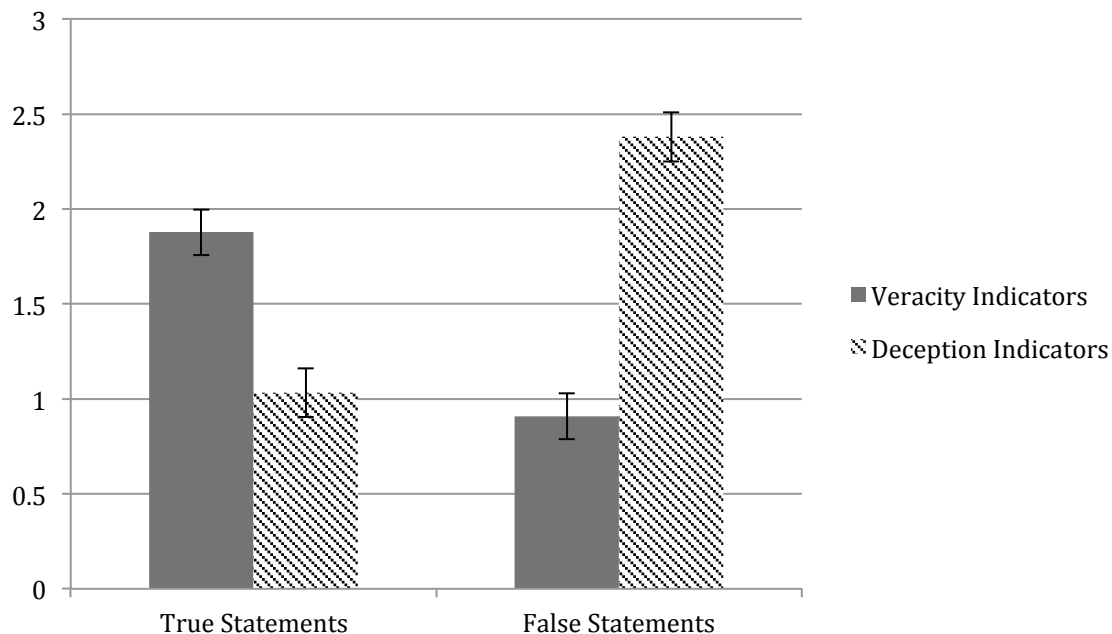
1. A repeated measures design for writing both true and false statements can work
2. A fixed order condition with participants writing true statements first and false statements second is methodologically problematic because of the presence of differential carry over effects; the content of the writing of the second false statement was influenced by the writing of the first true statement
3. A fixed order condition with participants writing false statements first and true statements second is methodologically better because of the lack of differential carry over effects; the content of the writing of the second true statement was not influenced by the writing of the first false statement

Year 2

In Year 2, participants from three language groups – English, Spanish, and Chinese – witnessed a video portraying an actual crime (from pilot studies in Year 1) and then wrote false and true statements about what they had witnessed in their respective languages. The statements were coded using various linguistic features of SA. The selected linguistic features discriminated between true and false witness statements and the effect sizes were relatively large (Figure 1). Importantly, language did *not* moderate the relationship between veracity and the coded features, indicating cross-language *similarity* in the efficacy of SA features to differentiate truths from lies.

Figure 1

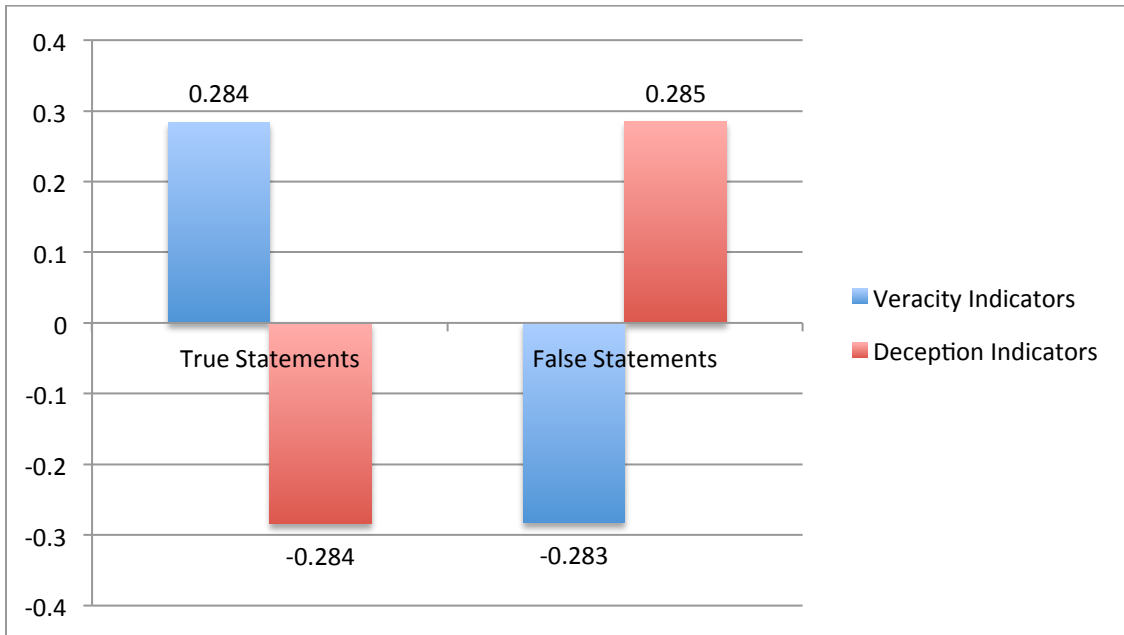
Differences between True and False Statements as Measured by Veracity and Deception Indicators (error bars refer to Standard Errors)



Year 3

In Year 3, participants from three language groups – English, Spanish, and Chinese – participated in a mock-crime experiment and then wrote false and true statements about what they had witnessed in their respective languages. The mock crime paradigm was chosen because it was substantially different than the eyewitness paradigm used in Year 2, and was much more personal. The statements produced by the participants were coded using various linguistic features of SA. The selected linguistic features discriminated between true and false witness statements and the effect sizes were relatively large (Figure 2). Importantly, language did *not* moderate the relationship between veracity and the coded features, indicating cross-language *similarity* in the efficacy of SA features to differentiate truths from lies. These findings replicated those in Year 2, and extended them to a substantially different research paradigm.

Figure 2

Residualized Means of Interaction of Veracity Conditions and Indicators

Cumulatively, these studies provided strong evidence that similar linguistic markers of veracity and deception can be used to differentiate truths from lies across languages.

Potential Impact or Translation to Military Applications

The positive results of this project provide Military, Intelligence and Law Enforcement professionals involved in source debriefing, interviewing, and interrogations with an important tool, and significantly enhance the cultural influence and intelligence-gathering capabilities of U.S. professionals who engage speakers from other cultures and languages. The findings enhance the effectiveness of information collectors who currently may rely solely or in large part on the practice of inducing detection apprehension in a source to seeking information by triggering nonverbal and paralinguistic cues that signal deception. Rather, the collector can focus on using structural and grammatical devices when it is presumed or believed that the individual is withholding information. Moreover seasoned investigators can learn to utilize both verbal and nonverbal analysis to assess information credibility. Because the findings from the project indicate that language does *not* moderate the ability of linguistic indicators to differentiate veracity from deception, there is good reason to believe that the techniques can generalize to *any* language. Given that, U.S. intelligence-gatherers and other military personnel, as well as federal, state, and local law enforcement, have another useful and viable tool at their disposal.

Performance Metrics

- Three manuscripts published in scientific, peer-reviewed publications
- Two manuscripts published in transitional publications for specific user communities (FBI Law Enforcement Bulletin and Journal of Tactics and Preparedness)
- Two conference presentations

Bibliography

Scientific Publications

- 1) Hwang, H. C., & Matsumoto, D. (in preparation). Linguistic cues of deception across multiple language groups.
- 2) Matsumoto, D., & Hwang, H. C. (in press). Emotional reactions to crime across cultures. *International Journal of Psychology*.
- 3) Matsumoto, D., Hwang, H. C., & Sandoval, V. A. (in press). Cross-language applicability of linguistic features associated with veracity and deception. *Journal of Police and Criminal Psychology*.

Transitional Publications

- 4) Sandoval, V. A., Matsumoto, D., Hwang, H. C., & Skinner, L. G. (in press). Exploiting verbal markers of deception across ethnic lines: An investigative tool for cross-cultural interviewing. *Federal Bureau of Investigation Law Enforcement Bulletin*.
- 5) Hwang, H. C., & Matsumoto, D. (2014). Detecting deception: Are you being lied to? *Tactics and Preparedness*, April(6), 6-7.

Conference Presentations

- 6) Matsumoto, D., & Hwang, H. C. (2014, September). *Cross-language applicability of statement analysis in detecting deception*. Paper presented at the 2014 Annual Conference of the Society for Police and Criminal Psychology, Las Vegas, NV.
- 7) Hwang, H. C., & Matsumoto, D. (2014, September). *Emotional reactions to crime*. Paper presented at the 2014 Annual Conference of the Society for Police and Criminal Psychology, Las Vegas, NV.

Appendices

1. Report on Pilot Study 1, Year 1 (in press at the *International Journal of Psychology*)
2. Report on main Study, Year 2 (in press at *Journal of Police and Criminal Psychology*)
3. Report on Main Study, Year 3 (currently submitted for publication)

**Appendix 1 – Report on Pilot Study 1, Year 1 (in press at the
International Journal of Psychology)**

In press, *International Journal of Psychology*

Running Head: Emotional reactions to crime

Emotional Reactions to Crime across Cultures

David Matsumoto and Hyisung C. Hwang

San Francisco State University and Humintell, LLC

All correspondence concerning this ms. should be sent to:

David Matsumoto

Department of Psychology

San Francisco State University

1600 Holloway Avenue

San Francisco, CA, 94132, USA

TEL: +1-510-338-1114

FAX: +1-510-217-9608

Email: dm@sfsu.edu

Abstract

Information about the emotions experienced by observers when they witness crimes would have important theoretical and practical implications, but to date no study has broadly assessed such emotional reactions. This study addressed this gap in the literature. Observers in seven countries viewed seven videos portraying actual crimes and rated their emotional reactions to each using 14 emotion scales. Observers reported significantly high levels of negative emotions including anger, contempt, disgust, fear, and sadness-related emotions, and anger, contempt, and disgust were the most salient emotions experienced by viewers across all countries. Witnesses also reported significantly high levels of positive emotions as well (compared to not feeling the emotion at all), which was unexpected. Country moderated the emotion ratings; post-hoc analyses indicated that masculine-oriented cultures reported less nervousness, surprise, excitement, fear, and embarrassment than feminine cultures.

Emotional Reactions to Crime across Cultures

The effects of mood on memory is a topic of long-standing inquiry (Bower, 1981; Bower, et al., 1994), and one arena in which such studies occur today concerns the effects of emotions on eyewitness testimony and witness credibility. A number of studies, for instance, have demonstrated that individuals who witness a negative emotional event may have enhanced memory for the central details of the event but impaired memory for peripheral details (Reisberg & Heuer, 2004; Safer, Christianson, Autry, & Osterlund, 1998). Houston et al. (2013) showed observers either an emotional (mugging) or neutral (conversation) scenario and obtained eyewitness recall from them about the perpetrator, critical incident, and environmental details. Emotionality improved completeness of perpetrator descriptions in a memory retrieval task but also impaired recognition of the perpetrators in a subsequent photo lineup. (The authors suggested that emotionality had differential effects on attending to “central” vs. “peripheral” details of an event.) Relatedly, several studies have also demonstrated a link between the consistency or inconsistency of the emotions displayed by victims when they recount their stories and judgments of their credibility (Dahl, et al., 2007; Kaufmann, Drevland, Wessel, Overskeid, & Magnussen, 2003). And there is a small but growing literature examining the effects of mood and emotion on juror’s processing and judgments (e.g., Semmler & Brewer, 2002).

That witnessing a crime should produce strong emotions is not surprising, given that crimes themselves are not merely acts void of feeling, but are replete with a myriad of emotions, including anger, fear, disgust, and even excitement and exhilaration (Canter & Ioannou, 2004; Canter, Kaouri, & Ioannou, 2003). Even what is known as cold or predatory aggression may not be entirely without emotion, as previously thought (Bushman & Anderson, 2001; Matsumoto & Hwang, in press). It would not be surprising, therefore, that observers who witnessed a crime also felt strong emotions, as they did in Houston, et al. (2013).

The question raised in this paper concerns exactly which emotions are aroused when individuals witness a crime. Of the variety of emotions that individuals can report, it is not clear that the field has a good grasp on exactly what kinds of emotions are elicited when observers witness a crime. Houston, et al. (2013) assessed irritation, annoyance, outrage, anger, happiness, sadness, sympathy, disgust, upset, fright, anxiety, relief, and nothing, and reported elevations of sympathy, disgust, annoyance, irritation, anger, sadness, outrage, and upset. Clearly these were sufficient to confirm the existence of “negative” emotions, which was the goal of that study. But that assessment may or may not have provided an accurate picture of the types of emotions elicited when people witness crimes. For example, it may be argued that annoyance, irritation, anger, outrage, and upset are synonyms of each other that assess essentially the same qualitative emotional state. Some may also argue about whether or not “sympathy” is an emotion.

Research elucidating more specifically the types of emotional experiences observers have when witnessing crimes may have important theoretical and practical ramifications. For example, if memories are mood-dependent, then information about the specific types of emotions elicited when observing crimes may suggest different effects on eyewitness recall for different types of emotions. Future research on this topic would be enhanced by targeting more specific emotional states rather than general “negative

emotions.” Such emotion-moderated effects would have practical implications for eyewitness testimony and juror processing.

One area of research that should inform this issue is the long history of scientific inquiry into the relationship between emotion and judgments of morality, because many studies have suggested that emotional reactions play an important role in mediating judgments of ethics and morality across cultures (Rozin & Fallon, 1987; Rozin, Lowery, Imada, & Haidt, 1999; Tangney & Fischer, 1995). In any society, crimes are acts that cross the boundaries of ethics and morality, break rules of social transgression, and defy social norms. Crimes are harmful not only to individuals but also to the community or state, and are a public wrong and forbidden and punishable by laws or social norms. Thus, witnessing crimes should elicit strong emotional reactions.

Within this area of research, emotions such as shame and guilt have received much attention as moral emotions (Shweder & Haidt, 2000; Tangney & Fischer, 1995). Additionally, recent work has focused on the emotions of anger, contempt and disgust and their relationship with ethics and morality. Rozin, et al. (1999) proposed that these emotions are often elicited by violations of three different types of moral codes originally proposed by Shweder and colleagues (Shweder, Much, Mahapatra, & Park, 1997). According to Rozin, et al. (1999), anger is linked to violations of individual rights and autonomy, contempt to violations of communal codes and hierarchy, and disgust to violations of purity and sanctity. Across four studies they showed that individuals in different cultures associated these emotions with specific examples of events that operationalized the proposed types of violations, with ratings of moral ethics violated by different types of situations, and with facial expressions of these emotions (Biehl, et al., 1997). Participants in their studies also produced distinct facial expressions of anger, contempt and disgust when reacting to violations of autonomy, community and divinity, respectively.

Hutcherson and Gross (2011) provided additional evidence that anger, contempt, and disgust are associated with moral judgments. Three of their five studies demonstrated that these emotions were associated with different types of antecedent appraisals related to ethics violations and morality, with anger evoked by appraisals of self-relevance, contempt by judgments of other’s incompetence or lack of intelligence, and disgust by appraisals that others are morally untrustworthy. Two studies demonstrated that individuals differentiated these emotions in terms of beliefs about their social consequences; individuals strongly preferred anger to contempt and disgust, and each emotion was associated with unique response profiles and judgments of real-life events.

Recent studies have also suggested that the combination of anger, contempt and disgust is what fuels terrorist acts and political aggression, acts that transgress moral and ethical boundaries (Matsumoto, Hwang, & Frank, 2013a, 2013b, 2014). These studies examined the emotions expressed by leaders of ideologically motivated groups that subsequently committed either an act of aggression or an act of non-violent resistance against an opponent outgroup. Speeches of leaders as they talked about their opponent outgroups were obtained at three points in time leading to an identified act and emotions expressed in those speeches were examined both verbally and nonverbally. The source materials analyzed in these studies spanned many different cultures and time periods. Leaders of groups that eventually committed acts of aggression expressed more anger,

contempt and disgust toward their opponent outgroups; leaders of groups that engaged in non-violent resistance did not differ in their expressions of these emotions.

The studies described above make a strong case that anger, contempt and disgust serve a special function vis-à-vis ethics and morality, and they do so similarly across cultures. If emotional reactions play an important role in judgments of ethics and morality across cultures, and if anger, contempt and disgust are emotions related to morality and ethics, then these same emotions should be especially salient when criminal acts are viewed because criminal acts themselves are transgressions of laws of ethics and rules of morality in a culture. We posit, therefore, that witnesses will experience anger, contempt, and disgust, and that these emotions will be the most salient emotions experienced.

But other emotions are also likely to be activated. When witnessing a crime, observers may feel threatened by the act or the perpetrator and fear for their own safety, either at that moment or later. Thus we would expect that observers experience fear-based emotions such as being scared, anxious, nervous, worried, or horrified. Observers may also empathize with the victims of crime and experience sadness, concern, anguish, or grief. Or they may lament the society in which they live and feel remorse or regret. For these reasons, we posit that observers would also feel negative emotions other than anger, contempt, and disgust, but that these are not as salient.

We have little reason to believe that observers would feel positive emotions when witnessing a crime (regardless of the fact that criminals themselves may feel positive emotions when committing crimes). And there is little reason to believe that the predictions described above will be moderated by country or culture, as all of the research described above documenting the relationship between emotion and judgments of morality and ethics have demonstrated similar effects across very disparate countries. This makes sense, as emotions are universal phenomena, and people of all cultures experience the same set of basic emotions regardless of race, culture, ethnicity, or nationality (Ekman, 1999; Izard, 2007; Matsumoto & Hwang, 2012). Although there are likely to be cultural differences in the absolute levels to which emotions are experienced, there is little reason to believe that witnessing a crime would not elicit negative reactions such as anger, contempt, disgust, fear, and sadness, or that anger, contempt, and disgust would not be the most salient.

This study addressed this gap in the literature. Participants in seven countries sampled by convenience viewed videos portraying actual crimes and rated their emotional reactions. We tested the following hypotheses, centered on the following research questions:

1. Which emotions do observers experience when witnessing a crime?
 - a. Hypothesis 1a: Witnesses will report significantly elevated (i.e., non-zero) levels of anger, contempt, and disgust.
 - b. Hypothesis 1b: Witnesses will also report significantly elevated (i.e., non-zero) levels of fear- and sadness-related emotions.
2. Which emotions are most salient?
 - a. Hypothesis 2: Anger, contempt and disgust will be more salient – i.e., have higher mean ratings – than other emotions when perceiving criminal acts.

We do not offer a prediction about the elevation of positive emotions because there is no reason to believe they would be significantly non-zero. And because the relationship between emotion and judgments of morality and ethics described above have occurred across cultures, we do not offer a prediction about the cultural moderation of the hypotheses, as we believe the emotion differences described above will occur across countries.

Methods

Stimuli

We searched the Internet for open source videos of actual crimes in different cultures. Surprisingly we found many such videos, many of which were posted by local police departments requesting the aid of the public in identifying persons of interest in the videos. Different types of crimes were represented including animal cruelty, armed robbery, arson, assault and battery, ATM theft, auto theft, burglary, hit and run, kidnapping, mugging, shooting, police brutality, shoplifting, pick pocketing, and vandalism. Our search resulted in obtaining an initial pool of 371 videos.

We then excluded videos that included any language in the video – either audio or written (subtitles) – because such commentary may have biased observers' reactions. We also excluded videos that were part of news reports (thus moderated by a newscaster) or that had technical difficulties (e.g., extremely low resolution). This resulted in a smaller pool of 94 videos from the U.S. or England, 48 from China, 6 from the Middle East, and 10 from Central or South Asia.

Although all videos were identified as “crime videos,” in many cases it was not clear that a crime had been committed unless the viewer had background information about the action in the video. For example a video of an auto theft of a person unlocking a car and driving off may seem innocuous unless the viewer knows that the driver is not the owner of the car. Because it was important to use videos that were clear that a crime was committed just by the observation of the contents of the video and not requiring any such background information or assumptions, two coders coded whether a crime had clearly been committed on each of the videos using a 5-point scale labeled 1, not clear at all, to 5, very clear.

Additionally we wanted to use videos that were relatively balanced in the amount of time devoted to the portrayal of the incident and before (prologue) and after (epilogue) the incident. An “incident” was defined as the act or event when the individual's behavior in that situation deviated from the norm. For this reason we also had coders log the time from the start of the video that the incident occurred and when the incident ended. Knowing these video times allowed us to calculate the amount of video times dedicated to the prologue, incident, and epilogue.

Videos were then selected for use in the study if the video had a crime rating of 5 from both coders, and the percentage of the video dedicated to the prologue and incident was each at least 30% of the entire length of each video. This resulted in the final selection of seven videos (country of origin of the video in parentheses):

- Video 1: Guy breaks into a car (China)
- Video 2: A woman shoplifts in a beauty supply store (U.S.)

- Video 3: A woman gets caught stealing from a store (U.S.)
- Video 4: Bangalore hit and run accident on the highway (India)
- Video 5: Guy throws brick into a car (England)
- Video 6: Burger King robbery at gunpoint (U.S.)
- Video 7: Animal cruelty – dog gets beaten to death (China)

We also selected one video to use as practice (motorcycle theft) for observers prior to their observing and rating the seven target videos. Thus eight videos were used in the study.

Observer Participants

A convenience sample of 555 observers from the U.S. ($n = 63$, $M_{age} = 33.55$, $SD_{age} = 13.37$, $n_{females} = 31$), India ($n = 143$, $M_{age} = 30.68$, $SD_{age} = 9.54$, $n_{females} = 54$), Ecuador ($n = 34$, $M_{age} = 29.15$, $SD_{age} = 12.02$, $n_{females} = 15$), Mexico ($n = 44$, $M_{age} = 29.80$, $SD_{age} = 9.97$, $n_{females} = 27$), Bolivia ($n = 30$, $M_{age} = 29.17$, $SD_{age} = 10.32$, $n_{females} = 19$), China ($n = 209$, $M_{age} = 22.98$, $SD_{age} = 8.61$, $n_{females} = 165$), and South Korea ($n = 32$, $M_{age} = 26.06$, $SD_{age} = 11.39$, $n_{females} = 23$) participated. They all self-reported as being born and raised their respective country and their first language corresponded to the language of their country. Local assistants recruited all observers from Ecuador, Mexico, Bolivia, China and South Korea in country; the U.S. Americans participated in our laboratory in Berkeley, California. The Indians were recruited using Amazon Mechanical Turk.

Judgment Tasks and Procedures

All survey materials were presented online and participants were provided the following instructions:

“The information gathered will be used for research examining cultural differences in perceptions of criminal acts. You will view several video scenes of acts, such as shoplifting, theft, etc. After each video, you will be asked some very basic questions about your thoughts about what you saw, such as ratings of believability, realism, probability of actual occurrence in your culture, the meaning of the act and its perceived legality, whether you have actually witnessed such an act in the past or heard about an actual event. You will also be asked basic demographic questions such as age, ethnicity and language. You will NOT be asked your name anywhere.”

After providing implied consent, participants were then shown the practice video. They were told to click the play button when ready, that they can enlarge to full screen by clicking the box [] at the bottom right of the video box, and to click ESC to return when done viewing.

After the video played, they were asked to rate how the video made them feel by indicating the extent to which they were currently experiencing any or all of the following emotions on a scale labeled 0, did NOT feel ANY of that emotion, to 8, an extreme amount of that emotion: Guilt, Fear, Anger, Embarrassment, Worry, Contempt, Excitement, Disgust, Amusement, Nervousness, Surprise, Interest, Sadness, and Pride. These emotion categories were selected in order to assess a broad range of qualitatively different emotional states, including qualitatively different positive and negative emotions.

Observers then rated the videos on 11 questions related to their beliefs about the crime. These ratings were not germane to this study and will not be mentioned further.

After completing the ratings, observers were shown the actual videos used in the study and given the same instructions as above for the practice video. The videos were shown in the order described above, from Video 1 through 7, because we considered them to be ordered in terms of emotional intensity, from least to most. We reckoned that ordering them in this fashion minimized the impact of emotional videos influencing the ratings of subsequent videos.

After the completion of the ratings for all videos, participants provided basic demographic information. Completion of their demographics marked the end of their participation in the study.

Computation of Emotion Scales

To reduce the 14 emotion ratings to a more manageable number of variables we computed a principal component analysis on the emotion ratings summed across videos, first for the entire sample. The analysis produced an interpretable three-factor solution that accounted for 74.13% of the total variance. The first factor accounted for 48.61% of the total variance and included anger, contempt and disgust; we labeled this factor ANCODI. The second factor accounted for an additional 19.27% of the variance and included excitement, amusement, pride and interest; we labeled this factor Positive Emotion. The third factor accounted for an additional 6.25% of the total variance and included fear, embarrassment, worry, nervousness, surprise, sadness and guilt; we labeled this factor Anxiety. We then computed scale scores for each of the three factors; Cronbach's α s were high and acceptable for each (.92, .87 and .81, respectively).

To establish the cross-cultural equivalence of the scale scores we also computed the same analyses separately for each of the countries. The same factor structures were obtained. Reliability estimates were also acceptable for each of the scales separately for each country ($.95 > \alpha > .86$; $.91 > \alpha > .69$; and $.97 > \alpha > .77$, for Anxiety, Positive Emotion and ANCODI, respectively).

Results

Hypotheses 1a and 1b: Which Emotions do Observers Experience when Witnessing a Crime?

We computed descriptive statistics for each of the emotion scale scores, separately for each country (Table 1). To examine if the emotion scale scores were significantly greater than zero (i.e., compared to "not feeling the emotion at all"), we computed one sample t-tests on each.¹ As predicted, the ANCODI and Anxiety scale scores were all high and significantly non-zero for the entire sample as well as for each country individually, $4.25 > \text{Cohen's } d > 1.56$, and $3.33 > d > 1.78$, respectively. Thus as

¹ The use of one sample t-tests against a population mean of zero raises interesting questions concerning assumptions about the normality of the distribution of the population mean and its hypothetical sampling distribution, which may affect the validity of the t-test. Readers are cautioned to interpret the results with this caveat.

predicted witnesses experienced elevated levels of anger, contempt, and disgust as well as fear and sadness related emotions, supporting Hypotheses 1a and 1b.

Interestingly the means for Positive Emotions were also significantly non-zero for the entire sample as well as for each country individually, and were associated with substantial effect sizes, $3.28 > d > 1.45$. To decompose the unexpected positive emotion effect, we examined whether the individual emotion ratings for excitement, amusement, interest, and pride were significantly non-zero for the entire sample and for each country separately using one sample t-tests. All four emotions were significantly greater than zero and associated with large effect sizes, $2.83 > d > 1.16$, $2.13 > d > 1.17$, $2.34 > d > 1.27$, and $2.45 > d > 1.15$, respectively. Thus it was apparent that not only did the witnesses experience the intended negative emotions, but they also experienced non-trivial amounts of positive emotions while watching the crime videos as well.

Hypothesis 2: Which Emotions are Most Salient?

To examine differences among the emotions, we computed a mixed Country (7) by Video (8) by Scale Type (3) ANOVA on the emotion ratings with Scale Type treated as repeated measures. The Scale Type main effect was significant, $F(2, 6012) = 539.59$, $p = .000$, $\eta_p^2 = .52$, indicating that in general the emotion scale scores differed from each other. We followed this main effect by computing a set of orthogonal difference contrasts. ANCODI ($M = 5.56$, $SE = .08$) had significantly higher ratings than Anxiety ($M = 4.18$, $SE = .08$), $F(1, 554) = 486.71$, $p = .000$, $\eta_p^2 = .47$, while Positive Emotions ($M = 2.55$, $SE = .07$) had significantly lower ratings than the combined ANCODI and Anxiety, $F(1, 554) = 797.19$, $p = .000$, $\eta_p^2 = .59$. Thus Hypothesis 2 was supported. Note the sizable effect sizes.

The above interpretations were qualified by a significant Country by Scale Type interaction, $F(12, 1002) = 18.21$, $p = .000$, $\eta_p^2 = .18$. We computed the same difference contrasts among the emotions separately for each country. The same comparisons were significant for all countries (separate listing of F tests available from authors). Observers in all countries gave the ANCODI the highest ratings, followed by Anxiety and then Positive Emotions.

The interpretation of the Scale Type effect was also qualified by a significant three-way interaction. Examination of the Scale Type differences on each of the videos separately for each country indicated that the same differences (i.e., ANCODI > Anxiety > Positive Emotion) occurred on each video across all of the countries. Thus the significant three-way interaction referred to differences in degree not direction.

The above findings were based on scale scores derived from the factor analyses described earlier. It was entirely possible that the means for anger, contempt and disgust each individually were not higher than the means of the other emotions. To examine this possibility we also computed an Emotion Type (14) by Country (7) two-way mixed ANOVA using the original emotions rated (see Table 2). The Emotion Type main effect was significant, $F(13, 6877) = 282.98$, $p = .000$, $\eta_p^2 = .35$. Anger, contempt and disgust did indeed receive the highest mean ratings across countries ($M_s = 6.15$, 5.46 , and 5.54 , respectively). Of these the lowest rated emotion (contempt) was still significantly higher than the next highest rated emotion (surprise), $F(1, 529) = 16.53$, $p = .000$, $\eta_p^2 = .03$.

Thus the findings reported using the scale scores derived from the factor analyses represented findings using the individual emotions rated as well.

Post-Hoc Analyses

The overall ANOVAs reported above also produced a significant Country main effect, $F(6, 501) = 19.62, p = .000, \eta_p^2 = .19$, as well as the interactions reported earlier. To examine potential cultural differences in the emotion ratings, we computed country-level rank order correlations between the emotion marginal means and each country's score on Hofstede's (2001) five cultural dimensions – Individualism, Power Distance, Masculinity, Uncertainty Avoidance, and Long-Term Orientation. (Cultural dimension data for Bolivia did not exist.) The only cultural dimension that approached significance was Masculinity, $\rho(6) = -.771, p = .072$, indicating that masculine-oriented cultures reported less emotions overall compared to feminine-oriented cultures. Rank order correlations separate for each emotion also indicated that Masculinity was negatively correlated with nervousness, surprise, excitement, fear, and embarrassment, $\rho(6) = -.925, p = .008$; $\rho(6) = -.882, p = .020$; $\rho(6) = -.924, p = .009$; $\rho(6) = -.772, p = .072$; and $\rho(6) = -.767, p = .075$, respectively.²

Discussion

As predicted witnesses reported significant amounts of different types of negative emotions when viewing the crimes. Of these, anger, contempt and disgust were the most salient emotions experienced by viewers across all six countries. Although country moderated the differences in the emotion effects, these referred to differences in the degree of difference, not direction, as anger, contempt and disgust had the highest means in each of the countries sampled, followed by fear and sadness-related emotions, and then positive emotions. Witnesses also reported significantly non-zero levels of positive emotions as well, which was unexpected.

These findings were not produced without limitations, one of which concerned the samples. On one hand because they were solely a convenience sample they did not represent a systematically chosen range of countries with which to test cultural differences and this may have contributed to our relative lack of country differences. On the other hand the countries that were sampled represented distinct world regions and a broad range on standard cultural values scales (Hofstede, 2001; Schwartz & Bardi, 2001). In any case readers should be cautioned in interpreting the country differences vis-à-vis the limitations in the sampling.

Another limitation concerned the sampling of the crime videos. Although we started with a fairly large pool of potential videos to use, we ultimately used a limited set of videos. They certainly did not represent the gamut of the types of crimes that occur in most societies nor did they reflect culture-specific crimes. Thus the findings reported

² Because of the different sex ratios in the different countries, we recomputed the post-hoc analyses separately for males and females. With only one exception (the correlation between Masculinity and fear for males), all other rank order correlations were high and negative ($-.60 < \rho < -1.00$).

above were limited to the crimes presented and it was entirely possible that different types of crimes may have produced different emotional reactions.

Another limitation concerned the sampling of emotions. Although the emotions we used assessed a wide range of affective experiences, it was possible that the assessment of other emotions may have produced different results. For example we did not assess shame, which has been linked to morality (Tangney & Fischer, 1995), although we did assess embarrassment. Future studies utilizing different emotion scales and different methods to assess emotions (e.g., facial expressions, physiological responses) may produce different results than what we report.

A final limitation concerned the laboratory-based nature of the data collection procedures. Witnessing a crime committed by strangers on strangers on a video played on a computer screen in a laboratory setting is very different than witnessing a crime in real life involving people one knows. In real life such an observation would involve raw emotions and sensory details like smells or sounds (Johnson, 1988; Johnson & Raye, 1981). Witnessing crimes replayed on video in the sterile confines of a laboratory undoubtedly influenced the nature of the emotions experienced, and future studies should examine emotional reactions in more real-life settings.

Regardless of these limitations, the findings contributed to the literature in several ways. First these data document the very diverse kinds of emotional experiences witnesses of crimes may experience, and as such may give clues into the minds of witnesses and the nature of memory recall. In particular, the data suggest that observers experience several qualitatively different types of “negative emotions” when witnessing a crime, and surprisingly positive emotions as well. Reporting high levels of anger, contempt, and disgust was not surprising, given previous research on the relationship between these emotions and judgments of violations of ethics and morality. As mentioned earlier these emotions have received special attention in terms of their socio-moral functions (Hutcherson & Gross, 2011; Matsumoto, et al., 2013a, 2013b, 2014; Rozin, et al., 1999). The current findings lend further credence to the notion that these emotions are especially important in understanding ethical and moral transgressions. They are the most salient emotions elicited when viewing ethical transgressions represented by crimes, thereby providing an emotional basis by which judgments of those transgressions are made.

That witnesses also reported experiencing elevated levels of fear and sadness-related emotions, including guilt, embarrassment, worry, and nervousness, was also not surprising, but are new to the field. These emotions are elicited by appraisals of threat and loss (Lazarus, 1991), and suggest that when witnessing a crime, observers may feel threatened by the act or criminal and fear for their own safety, either at that moment or later. Witnesses may have also empathized with the victims of the crimes and thus experienced sadness, concern, anguish, or grief.

That witnesses experienced significant, non-zero levels of positive emotions was unexpected and also new to the field. To be sure, these emotions may have occurred because of the laboratory nature of the task, and participants may have approached the task much like a video game. Or the elevation of these emotions may have reflected something about the human mind and its intrinsic interest for watching bad things happen

to others, perhaps reflecting a *shadenfreude* type of response. Future studies will need to examine this interesting twist on the findings.

We did not predict, nor did we find, that country moderated the differences among the emotions because previous research has demonstrated cross-cultural similarities in these emotional reactions vis-à-vis moral transgressions (Hutcherson & Gross, 2011; Rozin, et al., 1999). But the post-hoc analyses did indicate that masculine-oriented cultures reported less emotion overall, and less nervousness, surprise, excitement, fear, and embarrassment than feminine cultures. This finding was unexpected. Masculine cultures are those that value achievement, heroism, assertiveness, and material rewards for success, while feminine cultures value cooperation, modesty, caring for the weak, and quality of life (Hofstede, 2001). It is possible that the emotional reactions of the observers in more masculine cultures were tied into some sort of *machismo*, hero orientation in which ratings of fear-based emotions were deamplified. Future research examining a wider range of cultures on this dimension will need to examine this possibility more thoroughly.

The findings from this study suggest that observers may experience many different emotions when witnessing crimes, especially qualitatively different types of “negative emotions,” and perhaps also some types of positive emotions. These findings provide greater insights into the minds of eyewitnesses, and imply that further research differentiate the effects of qualitatively different types of elicited emotions on memory rather than classifying them in the general categories of “positive” or “negative.” Because mood congruent effects on memory has a long documented history in the literature, the present findings suggest that different emotional experiences will lead to different effects on memory in eyewitness recall and juror processing.

Understanding the emotions experienced by witnesses of crime also has implications for practitioners. Knowing the range of emotions that witnesses may report, and knowing that witnesses can and should report emotions in their statements about what they witnessed, should help practitioners in ferreting out true and false witness reports. Knowing exactly which emotions were experienced would provide investigators with a bridge to access memories, because mood binds memories and recall is easier when mood during recall is consistent with the mood during memory encoding (Bower, 1981).

Future research will need to examine emotional reactions to a wider range of moral and ethical transgressions in a wider range of cultures systematically chosen to operationalize relevant cultural dimensions, with special consideration to masculinity. Future research will also need to examine more carefully the role and function of anger, contempt and disgust in perceptions of crimes and other transgressions, and how these emotional reactions may be related to attitudes, values, and beliefs about laws, punishment and social norms in general. It may very well be the case that anger, contempt and disgust play a special role in the establishment and maintenance of cultural norms, and that one function of norms and punishments is to ameliorate levels of anger, contempt and disgust among group members to prevent retribution and social chaos. Such research will serve to contribute to knowledge about the sociocultural functions of emotions in general and about anger, contempt and disgust in particular. Finally, future

studies will need to delve into the interesting question of whether different types of emotional experiences produce differential effects on memory recall in eyewitnesses.

Author Notes

David Matsumoto and Hyisung C. Hwang, Department of Psychology, San Francisco State University and Humintell, LLC.

The authors would like to thank Jeff Fujimoto for his aid in collecting data in Ecuador, Tony Sandoval for his aid in collecting data in Bolivia, and Xiaohang Feng for her aid in collecting data in China. This report was prepared with the support of research grant FA9550-11-1-0306 from the U.S. Air Force Office of Scientific Research.

Correspondence concerning this article should be addressed to David Matsumoto, Department of Psychology, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA, 94132, telephone (415) 338-1114, fax (510) 217-9608, or e-mail dm@sfsu.edu.

References

- Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., & Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability Data and Cross-National Differences. *Journal of Nonverbal Behavior*, 21, 3-21.
- Bower, G. H. (1981). Mood and memory. *American Psychologist*, 36, 129-148.
- Bower, G. H., Lazarus, R., LeDoux, J. E., Panksepp, J., Davidson, R. J., & Ekman, P. (1994). What is the relation between emotion and memory? In R. J. D. Paul Ekman (Ed.), *The nature of emotion: Fundamental questions. Series in affective science*. (pp. 301-318): Oxford University Press, New York, NY, US.
- Bushman, B. J., & Anderson, C. A. (2001). Is it time to pull the plug on the hostile versus instrumental aggression dichotomy? *Psychological Review*, 108, 273-279.
- Canter, D. V., & Ioannou, M. (2004). Criminals' emotional experiences during crimes. *International Journal of Forensic Psychology*, 1, 71-81.
- Canter, D. V., Kaouri, C., & Ioannou, M. (2003). The facet structure of criminal narratives. In S. Levy & D. Elizur (Eds.), *Facet theory: Towards a cumulative social science* (pp. 27-38). Ljubljana, Slovenia: Faculty of Arts, Center for Educational Development.
- Dahl, J., Enemo, I., Drevland, G. C. B., Wessel, E., Eilertsen, D. E., & Magnussen, S. (2007). Displayed emotions and witness credibility: A comparison of judgements by individuals and mock juries. *Applied Cognitive Psychology*, 21, 1145-1155.
- Ekman, P. (1999). Basic emotions. In T. Dalgleish & T. Power (Eds.), *The handbook of cognition and emotion* (pp. 45-60). Sussex, United Kingdom: John Wiley and Sons, Ltd.
- Hofstede, G. H. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Houston, K. A., Clifford, B. R., Phillips, L. H., & Memon, A. (2013). The emotional eyewitness: The effects of emotion on specific aspects of eyewitness recall and recognition performance. *Emotion*, 13, 118-128.
- Hutcherson, C. A., & Gross, J. J. (2011). The moral emotions: A social-functionalist account of anger, disgust, and contempt. *Journal of Personality and Social Psychology*, 100, 719-737.
- Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, 2, 260-280.
- Johnson, M. K. (1988). Reality monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General*, 117, 390-394.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.

- Kaufmann, G., Drevland, G. C. B., Wessel, E., Overskeid, G., & Magnussen, S. (2003). The importance of being earnest: Displayed emotions and witness credibility. *Applied Cognitive Psychology, 17*, 21-34.
- Lazarus, R. (1991). *Emotion and adaptation*. New York: Oxford University Press.
- Matsumoto, D., & Hwang, H. C. (in press). Facial signs of imminent aggression. *Journal of Threat Assessment and Management*.
- Matsumoto, D., Hwang, H. C., & Frank, M. G. (2013a). Emotional language and political aggression. *Journal of Language and Social Psychology, 32*, 452-468.
- Matsumoto, D., Hwang, H. C., & Frank, M. G. (2013b). Emotions expressed by leaders in videos predict political aggression. *Behavioral Sciences of Terrorism and Political Aggression*. doi: DOI:10.1080/19434472.2013.769116
- Matsumoto, D., Hwang, H. C., & Frank, M. G. (2014). Emotions expressed in speeches by leaders of ideologically motivated groups predict aggression. *Behavioral Sciences of Terrorism and Political Aggression, 6*, 1-18. doi: 10.1080/19434472.2012.716449
- Matsumoto, D., & Hwang, H. S. (2012). Culture and emotion: The integration of biological and cultural contributions. *Journal of Cross-Cultural Psychology, 43*, 91-118.
- Reisberg, D., & Heuer, F. (2004). Memory for emotional events. In D. Reisberg & P. Hertel (Eds.), *Memory and emotion* (pp. 1-41). New York, NY: Oxford University Press.
- Rozin, P., & Fallon, A. E. (1987). A perspective on disgust. *Psychological Review, 94*, 23-41.
- Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 75*, 574-585.
- Safer, M. A., Christianson, S., Autry, M. A., & Osterlund, K. (1998). Tunnel memory for traumatic events. *Applied Cognitive Psychology, 12*, 99-117.
- Schwartz, S. H., & Bardi, A. (2001). Value hierarchies across cultures: Taking a similarities perspective. *Journal of Cross-Cultural Psychology, 32*, 268-290.
- Semmler, C., & Brewer, N. (2002). Effects of mood and emotion on juror processing and judgments. *Behavioral Sciences and the Law, 20*, 423-426.
- Shweder, R. A., & Haidt, J. (2000). The cultural psychology of the emotions: Ancient and new. . In M. Lewis & J. M. Haviland (Eds.), *The handbook of emotions* (pp. 397-414). New York: Guilford Press.
- Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "Big Three" of morality (autonomy, community, divinity) and the "Big Three" explanations of suffering. In B. A. & P. Rozin (Eds.), *Morality and health* (pp. 119-169). New York: Routledge.

Tangney, J., & Fischer, K. W. (Eds.). (1995). *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride*. New York: Guilford Press.

Table 1. *Means and Standard Errors for the Emotion Scales separate by Country*

Country	ANCODI	Anxiety	Positive Emotions
U.S.	5.94 (.22)	4.08 (.21)	3.52 (.21)
India	5.31 (.14)	4.84 (.14)	3.89 (.17)
Ecuador	3.90 (.42)	3.35 (.32)	1.75 (.13)
Mexico	6.16 (.32)	4.27 (.28)	2.12 (.17)
Bolivia	6.56 (.37)	5.13 (.31)	2.23 (.17)
China	5.51 (.13)	3.63 (.12)	1.61 (.08)
South Korea	6.62 (.28)	5.16 (.27)	3.33 (.18)

Table 2

Descriptive Statistics (Means and Standard Errors) for each of the Emotions, Separately for each Country

Emotion	Country						
	U.S.	India	Ecuador	Mexico	Bolivia	China	South Korea
Anger	6.18 (.27)	5.82 (.18)	4.18 (.33)	6.76 (.29)	6.94 (.35)	6.42 (.13)	6.95 (.35)
Contempt	5.38 (.31)	5.23 (.21)	3.86 (.39)	6.08 (.34)	6.85 (.41)	3.90 (.16)	7.12 (.40)
Disgust	6.38 (.29)	5.35 (.20)	3.67 (.36)	5.65 (.32)	5.89 (.38)	6.23 (.15)	5.85 (.38)
Guilt	3.18 (.28)	4.54 (.19)	2.02 (.34)	2.29 (.30)	1.74 (.36)	2.20 (.14)	3.15 (.36)
Fear	4.31 (.29)	4.75 (.20)	3.43 (.36)	4.28 (.31)	4.72 (.38)	3.96 (.14)	5.22 (.37)
Embarrassment	3.27 (.30)	4.92 (.20)	3.20 (.37)	4.33 (.33)	5.08 (.40)	2.63 (.15)	5.72 (.39)
Worry	4.72 (.28)	5.39 (.19)	4.08 (.34)	6.01 (.30)	6.90 (.37)	4.55 (.14)	6.52 (.36)
Nervousness	4.22 (.29)	4.79 (.19)	3.53 (.35)	3.83 (.31)	5.09 (.37)	3.97 (.14)	5.58 (.37)
Surprise	4.72 (.29)	5.01 (.20)	3.56 (.36)	4.31 (.32)	6.15 (.38)	4.54 (.15)	6.51 (.38)
Sadness	4.43 (.29)	5.27 (.20)	3.59 (.36)	4.84 (.31)	6.21 (.38)	3.37 (.15)	3.61 (.37)
Excitement	3.40 (.23)	4.22 (.15)	1.61 (.28)	1.63 (.25)	1.72 (.30)	1.77 (.11)	5.65 (.29)
Amusement	3.28 (.20)	4.22 (.14)	1.66 (.25)	1.46 (.22)	1.30 (.27)	1.59 (.10)	1.72 (.26)
Interest	4.63	4.13	2.33	4.15	4.66	1.75	4.85

	(.26)	(.18)	(.32)	(.28)	(.34)	(.13)	(.34)
Pride	2.89	3.37	1.39	1.25	1.25	1.32	1.00
	(.20)	(.13)	(.24)	(.21)	(.26)	(.10)	(.25)

**Appendix 2 - Report on Main Study, Year 2 (in press at
Journal of Police and Criminal Psychology)**

Running Head: Cross-Language Applicability in Detecting Deception

In press, *Journal of Police and Criminal Psychology*

Cross-Language Applicability of Linguistic Features associated with Veracity and
Deception

David Matsumoto and Hyisung C. Hwang

San Francisco State University and Humintell, LLC

Vincent A. Sandoval

Federal Bureau of Investigation (retired)

All correspondence concerning this ms. should be sent to:

David Matsumoto

Department of Psychology

San Francisco State University

1600 Holloway Avenue

San Francisco, CA, 94132, USA

TEL: +1-510-338-1114

FAX: +1-510-217-9608

Email: dm@sfsu.edu

Author Notes

David Matsumoto and Hyisung C. Hwang, Department of Psychology, San Francisco State University and Humintell, LLC. Vincent A. Sandoval, Federal Bureau of Investigation (retired).

The authors would like to thank the following individuals for their aid in the collection of data in Study 1: Jeff Fujimoto for his aid in collecting data in Ecuador, Angel Avendano for his aid in collecting data in Bolivia, and Xiaohang Feng for her aid in collecting data in China. This report was prepared with the support of research grant FA9550-11-1-0306 from the U.S. Air Force Office of Scientific Research.

Correspondence concerning this article should be addressed to David Matsumoto, Department of Psychology, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA, 94132, telephone (415) 338-1114, fax (510) 217-9608, or e-mail dm@sfsu.edu.

Abstract

One technique for examining written statements or interview transcripts for verbal cues of veracity and lying involves the analysis of linguistic features and grammatical structures associated with word usage. This technique is commonly referred to as Statement Analysis (SA). There are varying degrees of empirical support for different SA techniques and for specific linguistic markers; what is less known in the literature is the degree to which verbal indicators of veracity and lying vary across languages. We examined this research question. Participants from three language groups – English, Spanish, and Chinese – witnessed a video portraying an actual crime and then wrote false and true statements about what they had witnessed in their respective languages. The statements were coded using various linguistic features of SA. The selected linguistic features discriminated between true and false witness statements and the effect sizes were relatively large. Importantly, language did *not* moderate the relationship between veracity and the coded features, indicating cross-language similarity in the efficacy of SA features to differentiate truths from lies.

Evidence for Cross-Language Applicability of Linguistic Features associated with Veracity and Lying

One technique for examining language for clues to lying involves the analysis of linguistic features and grammatical structures associated with word usage, commonly referred to as Statement Analysis (SA). SA is based on the premise that word use and grammar structures differ when people lie as opposed to when they tell the truth. Because words make up sentences and sentence construction follows a predetermined set of grammatical rules, a careful examination of word use and grammar structures should identify specific features that can help detect deception.

SA has its roots in psycholinguistic research in the early 1900s but has received more contemporary reception within forensic psychological and law enforcement circles as a result of the work of Undeutsch (1989) and a technique known as Statement Validity Analysis (SVA). SVA was founded on a hypothesis that statements based on actual memories differ from fabricated or fantasized statements (Undeutsch, 1989). The crucial parts of SVA involve a criteria-based content analysis (CBCA) and an evaluation of CBCA outcomes using a Validity Check-List with criteria organized around categories such as general, unusual, motivational and stylistic features. In addition to SVA a number of other techniques that involve the analysis of the grammatical structures to make inferences about deception and truthfulness have emerged, including Theoretical Verbal Analysis (TVA; Connelly, et al., 2006), Reality Monitoring (RM; Johnson & Raye, 1998), Scientific Content Analysis (SCAN; Sapir, 1996), and Investigative Discourse Analysis (IDA), which is an extension of CBCA and similar to SCAN (Rabon, 1994).

Research examining the efficacy of various SA techniques has provided evidence for many of them to detect truths from lies at better than chance accuracies (Porter & Yuille, 1996; Vrij, 2007; Vrij & Mann, 2006; Zaparniuk, Yuille, & Taylor, 1995). For example, CBCA has been linked to empirically based knowledge about naturalistic memory and to a fair amount of research demonstrating the validity of many of its criteria (Porter, Birt, Yuille, & Lehman, 2000; Porter & Yuille, 1996; Zaparniuk, et al., 1995). RM is also based on a solid empirical base of knowledge about memory (Johnson, 1988; Johnson & Raye, 1981) and reviews examining its usefulness in detecting deception have confirmed the validity of many of its criteria (Masip, Sporer, Garrido, & Herrero, 2005; Sporer, 2004).

What is less known in the literature is whether verbal indicators of veracity and lying vary across languages, and whether or not the same SA features can be used to differentiate truths from lies across languages, because most of the research to date has analyzed source materials produced in one language (usually English by native English speakers). To be sure there are studies examining the verbal indicators of veracity and lying in non-English languages (Masip, Bethencourt, Lucas, Sanchez-San Segundo, & Herrero, 2012; Ruby & Brigham, 1997; Schelleman-Offermans & Merckelbach, 2010). For instance, Masip, et al. (2012) asked Spanish students to write a truthful or deceptive story; subsequent analyses demonstrated that truthful and false stories differed on plausibility, details, consistency, emotions, and structure. Schelleman-Offermans and Merckelbach (2010) asked students in The Netherlands to write a true or false story (in Dutch) about an aversive situation in which they had been victims (e.g., of gossip, bullying, robbery, etc.). The statements were coded using nine CBCA criteria; three

differentiated true statements from false: logical structure, contextual embedding, and attribution of the perpetrator's state. Although these studies are suggestive of the potential cross-language applicability of SA, however, comparing results across these studies to make generalizations is problematic because study differences confound the languages examined and their results (but study differences also speak to the robustness of the findings).

We remedied this situation by analyzing the linguistic indicators of veracity and lying in a realistic, moderately high stakes scenario and by examining three very different languages within the same study. There are important theoretical reasons to investigate this issue. Cross-language consistency may provide evidence for potential pancultural similarity in the underlying psychological effects of lying and similarity as reflected in the linguistic choices that mark those effects. Such effects would suggest cross-cultural similarity in the structure of memory, the recall of information from memory, and the psychological demands placed on individuals who lie about that recall, and would point to a potential universal mechanism of lying that can be identified by specific linguistic markers. If the rules of grammar and deep structure of language (Chomsky, 1957, 1972) and the principles of memory and recall (Undeutsch, 1989) are similar across cultures, then verbal indicators of truths and lies may occur regardless of culture, ethnicity, and language.

In this study individuals from three language groups – English, Spanish, and Chinese – witnessed a crime and were asked to write false and true statements about what they had witnessed. Both statements were written in the native language of the participants. The three language groups were chosen to sample a broad range of cultural/linguistic differences that may influence indicators of veracity and deception. These languages also represented major language/cultural groups around the world, as well as the U.S.

Linguistic and Grammatical Markers of Veracity and Lying Used in this Study

There are some commonalities among the various SA techniques as they are based on a relatively common understanding of the nature of human memory and verbal recall of that knowledge. Differences among these systems occur concerning the specific linguistic categories considered indicative of veracity or lying and in the amount of scientific evidence that exists for all the various features of each system, especially across different languages and cultural/ethnic groups. Because SA techniques allow for the analysis of many different types of linguistic and grammatical markers with commonalities, and because of differences across studies in the degree of empirical support for specific categories within specific techniques, we selected for use in this study an eclectic group of SA categories from different techniques deemed most relevant for the source materials produced.

More specifically, for this study, we were not concerned with testing the applicability of any one SA technique across languages, but rather whether or not a core set of SA categories that exist across techniques was applicable across languages. Thus the categories we selected for use in this study were those that occurred across different SA techniques (e.g., CBCA and Reality Monitoring) and for which there was empirical support for use with people of different cultural/ethnic backgrounds. And, we were

interested in including SA categories that have proven to be operationally relevant in the field (i.e., categories that have also been the most effective in actual investigations based on the experiences of the third author). The following categories were selected on an a priori basis for use in this study before any statements were coded.

Indicators of veracity

Structural balance. In response to an open-ended question designed to elicit information, a writer will typically include information about what transpired prior to (prologue), during (incident), and after (epilogue) the incident (Johnson, 1988). The *prologue* provides contextual information relative to the setting such as details pertaining to time of day, place and the people involved. The *incident* pertains to that portion of the statement where the actual criminal event takes place and begins at that point in the narrative where an investigator would conclude a crime is taking place and warrant initiating an investigation (Rabon, 1994). The *epilogue* consists of subordinate information such as the writer's emotional reaction to the incident or efforts to contact law enforcement. Research and experience have demonstrated that a good indicator of veracity is balance within the structure of a written statement (Adams & Jarvis, 2006; Rabon, 1994), because there is an expectation that when writers discuss a specific event they will dedicate most of their statement to that event. Balance is determined by ascertaining how much space the writer dedicates to each of the three component parts of a statement. Researchers differ on the precise percentages that each component part should possess but all agree that the incident should be at least equal to or greater than the prologue and epilogue (Rabon, 1994; Sapir, 1996). When a statement contains an inordinately long prologue, that statement will often be a deceptive statement (Adams & Jarvis, 2006; Rabon, 1994; Rudacille, 1994; Sapir, 1996).

Word count. A number of studies have demonstrated that liars use fewer words than truth-tellers (DePaulo, et al., 2003; Newman, Pennebaker, Berry, & Richards, 2003). This is likely the result of liars using a strategy of simply omitting important details from their written statement.

Unique sensory detail and spatial detail (USD-SD). Researchers have postulated that there are identifiable differences between truthful and fabricated statements by identifying the presence of and location within the statement of specific types of details within those statements (Johnson, 1988; Johnson & Raye, 1981; Porter & Yuille, 1996; Undeutsch, 1989). Unique sensory detail (USD) pertains to specific descriptions generated by the five sensory perceptions (sight, sound, touch, smell, taste and touch). Spatial detail (SD) pertains to specific locations and the physical relationships of objects, people, etc., in relation to one another (Adams & Jarvis, 2006). The expectation is that truthful writers who discuss a specific event will provide requisite detail about that event. While evidence for these kinds of details to differentiate truths from lies comes from several theoretical frameworks, both CBCA (Porter & Yuille, 1996; Undeutsch, 1989; Vrij, 2007) and the Reality Monitoring frameworks (Johnson, 1988; Johnson & Raye, 1981) in particular have provided strong evidence to suggest that individuals who recall previously encoded events truthfully report more sensory and spatial details because these details are encoded in memory along with the factual content of the event.

Emotion. The presence of the writer's emotion or affective responses in reaction

to the incident, such as fear, anger, shock, or embarrassment can also differentiate truths from lies. Both CBCA (Porter & Yuille, 1996; Undeutsch, 1989; Vrij, 2007) and the Reality Monitoring frameworks (Johnson, 1988; Johnson & Raye, 1981) provide evidence to indicate that individuals who recall previously encoded events truthfully also report more emotions in relation to the event because emotions are encoded in memory when events occur. Other studies have suggested that this is especially true when the mentions of emotion are found in the epilogue of the written statement (Adams & Jarvis, 2006).

Indicators of lying

Extraneous information. A number of studies (DePaulo, et al., 2003; Matsumoto, Hwang, & Sandoval, 2013; Vrij, 2007) have demonstrated that truth tellers provide more details relevant to the question raised, whereas liars provide more information that is irrelevant, which we refer to as extraneous information. Extraneous information is information that does not answer the question posed, and may be used to justify the liars' actions, deflect the question because they may not want to respond to that specific question, help liars distance themselves from the act of lying or the content of the lie, or aid liars in exerting control over the interview (Adams, 1996). Matsumoto and colleagues (2013) reported that liars from different ethnic groups produced more extraneous information when writing statements in English.

Equivocation. Equivocation refers to information that is not relevant to the question that was posed to elicit the statement. Equivocation words qualify statements, allowing liars to distance themselves from the act or content of lying by tempering the action about to be described or by discounting the message even before it is transmitted (Weintraub, 1989). Equivocation consists of words or phrases such as “maybe”, “believe”, “kind of”, “sort of”, “about”, or “to the best of my knowledge”, which suggest that the interviewee is being intentionally vague or ambiguous. Matsumoto and colleagues (2013) reported that liars from different ethnic groups produced more equivocation when writing statements in English.

Non-prompted negation (NPN). When responding to a question such as, “Tell me what you did in the file room,” the expectation is that individuals will respond by providing information pertaining to what they actually did (Rudacille, 1994; Sapir, 1996; Weiner & Mehrabian, 1968). Therefore a response about what the individual did not do (e.g., “I did not see a car hit anyone”) does not answer the question and is an example of NPN. Negation in discourse or statements may be an indicator of deception inasmuch as respondents may use it to carefully omit their involvement in a crime (Adams & Jarvis, 2006), and there are generally more negative statements in deceptive oral narratives than in truthful oral accounts (Hauch, Blandon-Gitlin, Masip, & Sporer, 2012; Newman, et al., 2003; Porter, et al., 2000). Matsumoto and colleagues (2013) reported that liars from different ethnic groups produced more NPN both when writing statements in English and in oral interviews.

Moderating adverbs. We identify three types of moderating adverbs. (1) *Intensifying adverbs* such as “very,” “really,” “truthfully,” or “honestly” are typically used when a communicator is attempting to convince another person of something. (2) *Minimizing adverbs* such as “only,” “just,” “simply,” “merely” are typically used to

downplay or minimize the role of the actor, who is generally the communicator him or herself. (3) *Editing adverbs* such as “after,” “then,” “next,” “while,” “so,” “thereafter,” or “when” may indicate a temporal lacunae (Rabon, 1994; Schafer, 2007) suggesting that the communicator is intentionally editing information and as such, something that might be crucial to an inquiry may be missing from the discourse. Because lies of omission are more common than lies of commission, and because liars tend to use fewer words than truth tellers (DePaulo, et al., 2003), editing adverbs provide liars with a simple yet strategic means of telling the truth up to a certain point, omitting crucial information and then picking up again by telling the truth. Matsumoto and colleagues (2013) reported that liars from different ethnic groups produced more moderating adverbs both when writing statements and in oral interviews.

Passive voice. When describing their actions, people will generally assume responsibility for those actions by employing the active voice (i.e., the agent engaging in the action described by a verb is the subject of the sentence). For example, liars who attempt to conceal their identity as an actor, such as firing a pistol, may attempt to distance themselves from the action by employing the passive voice, “the pistol was fired.” Passive voice occurs when the object of an action verb appears as the subject of the sentence. It may be used when liars attempt to conceal their identity as an actor, distancing themselves from the action of the verb (Connelly, et al., 2006; Rudacille, 1994).

Overview of the Study

Participants from three language groups – English, Spanish, and Chinese – witnessed a crime and were asked to write false and true statements about what they had witnessed, in their native language. Participants were led to believe that their statements would be read by investigators who would make a determination about the believability of the statements, and that there were rewards and punishments for the participants depending on those determinations. The statements they produced were coded for the linguistic features of veracity and lying described above. We hypothesized that the coded features could differentiate true and false statements across the three languages.

Methods

Stimuli

Initial pool. Because the study involved participants writing about a crime they had witnessed, it was necessary to first conduct a pilot study to select a crime video that could be used. We wanted to use a video that depicted an actual crime, that aroused emotions in the part of witnesses/viewers (as would most crimes), and that allowed for a fair test of the structural features of true and false statements (vis-à-vis structural balance). Thus we conducted a pilot study in order to identify the video that would be used in the main study.

First, we searched the Internet for open source videos of actual crimes in different cultures. Surprisingly we found many such videos, many of which were posted by local police departments requesting the aid of the public in identifying persons of interest in the videos. Different types of crimes were represented including animal cruelty, armed robbery, arson, assault and battery, ATM theft, auto theft, burglary, hit and run,

kidnapping, mugging, murder, police brutality, shoplifting, pick pocketing, and vandalism. Our search resulted in obtaining an initial pool of 371 videos.

We then excluded videos that included any language in the video – either audio or written (subtitles) – because such commentary may have biased observers' reactions. We also excluded videos that were part of news reports (moderated by a newscaster) or that had technical difficulties (e.g., extremely low resolution). This resulted in a smaller pool of 94 videos from the U.S. or England, 48 from China, 6 from the Middle East, and 10 from Central or South Asia.

Although all videos were identified as “crime videos,” in many cases it was not clear that a crime had been committed unless the viewer had background information about the action in the video. For example, a video of an “auto theft” of a person unlocking a car and driving off is innocuous unless the viewer knows that the driver is not the owner of the car. Because it was important to use videos that were clear that a crime was committed just by the observation of the video and not requiring any such background information or assumptions, two coders coded whether a crime had clearly been committed on each of the videos using a 5-point scale labeled 1, not clear at all, to 5, very clear.

Additionally we wanted to use videos that were relatively balanced in the amount of time devoted to the portrayal of the incident and before (prologue) and after (epilogue). As described earlier, Structural Balance is one of the features that may differentiate true and false statements; thus we needed videos that were balanced themselves so as not to skew balance in the resulting statements produced. In the videos, an “incident” was defined as the act or event when the individual's behavior in that situation deviated from the norm. Thus we also had coders log the time from the start of the video that the incident occurred and when the incident ended. We then calculated the amount of video times dedicated to the prologue, incident, and epilogue.

Videos were selected for use in the study if the video had a crime rating of 5 from both coders and the percentage of the video dedicated to the prologue and incident was each at least 30% of the entire length of each video. This resulted in the final selection of seven potential videos (country of origin of the video in parentheses):

- Video 1: Guy breaks into a car (China)
- Video 2: A woman shoplifts in a beauty supply store (U.S.)
- Video 3: A woman gets caught stealing from a store (U.S.)
- Video 4: Bangalore hit and run accident on the highway (India)
- Video 5: Guy throws brick into a car (England)
- Video 6: Burger King robbery at gunpoint (U.S.)
- Video 7: Animal cruelty – dog gets beaten to death (China)

We also selected one video to use as practice (motorcycle theft) for observers prior to their observing and rating the seven target videos. Thus eight videos were rated.

Observers, judgment tasks, and procedures. A total of 555 observers from the U.S. ($n = 63$), India ($n = 143$), Ecuador ($n = 34$), Mexico ($n = 44$), Bolivia ($n = 30$), China ($n = 209$), and South Korea ($n = 32$) participated. They all self-reported as being born and raised in their respective country and their first language corresponded to the language of

their country. Local assistants recruited all observers from Ecuador, Mexico, Bolivia, China and South Korea in country; the U.S. Americans participated in our laboratory in Berkeley, California. The Indians were recruited using Amazon Mechanical Turk.

All survey materials were presented online and participants were provided the following instructions:

“The information gathered will be used for research examining cultural differences in perceptions of criminal acts. You will view several video scenes of acts, such as shoplifting, theft, etc. After each video, you will be asked some very basic questions about your thoughts about what you saw, such as ratings of believability, realism, probability of actual occurrence in your culture, the meaning of the act and its perceived legality, whether you have actually witnessed such an act in the past or heard about an actual event. You will also be asked basic demographic questions such as age, ethnicity and language. You will NOT be asked your name anywhere.”

After providing implied consent, participants were then shown the practice video. They were told to click the play button when ready, that they can enlarge to full screen by clicking the box [] at the bottom right of the video box, and to click ESC to return when done viewing. After the video played, they were asked to rate how the video made them feel by indicating the extent to which they were currently experiencing any or all of the following emotions on a scale labeled 0, did NOT feel ANY of that emotion, to 8, an extreme amount of that emotion: Guilt, Fear, Anger, Embarrassment, Worry, Contempt, Excitement, Disgust, Amusement, Nervousness, Surprise, Interest, Sadness, and Pride. They also completed a set of attitude and belief ratings not germane to this study.

After completing the ratings, observers were shown the actual videos used in the study and given the same instructions as above for the practice video. The videos were shown in the order listed above, from Video 1 through Video 7, because we considered them to be ordered in terms of emotional intensity, from least to most. We ordered them in this fashion to minimize the impact of emotional videos influencing the ratings of subsequent videos.

After the completion of the ratings for all videos, participants provided some basic demographic information. Completion of their demographics marked the end of their participation in the study.

Final selection. The goal of the analyses was to determine cultural differences or similarities in the ratings of the videos in order to select videos for use in the main study that were relatively cross-culturally invariant and elicited the greatest emotion. To elucidate this issue, we computed intraclass correlations (ICCs) across means of the 14 emotion ratings considering the seven countries as raters, separately for each video. These analyses allowed us to determine if the relative ranking among the 14 emotions was consistent across countries or not. ICCs can be computed in two ways, one using absolute agreement as a basis and a second way using consistency as a basis. We were particularly interested in ICCs based on absolute agreement, as these would indicate the degree to which the relative rankings of the 14 emotion means were similar across cultures *and* anchored to a similar absolute score.

There was considerably high agreement across the countries on the relative rankings across means of the 14 emotions rated for each of the seven videos and across all videos overall (Table 1). These findings suggested that there was a great deal of consistency across the countries in the means of their emotional profiles for each of the videos.

We then examined the overall marginal means of the 14 emotion ratings for each of the videos to determine which video(s) elicited the greatest overall amount of emotion. (Examination of the marginal means was justified given the high ICC values obtained in the results above.) The videos ranked in size of their emotion marginal means were Video 7 (Animal Cruelty, $M = 4.96$, $SE = .09$), Video 4 (Hit and Run, $M = 4.45$, $SE = .08$), Video 6 (Robbery, $M = 4.26$, $SE = .09$), Video 5 (Brick, $M = 4.07$, $SE = .09$), Video 2 (Shoplifting, $M = 3.84$, $SE = .08$), Video 1 (Car, $M = 3.83$, $SE = .08$), and Video 3 (Stealing, $M = 3.82$, $SE = .08$). Because we wanted to use the video that elicited the highest overall emotion ratings, we initially attempted to obtain IRB approval for use of Video 7 (Animal Cruelty). Unfortunately approval was not obtained for that video; thus Video 4 (Hit and Run) was used in this study.

Participants

The participants in the main study included 43 Chinese (20 females, 22 males, 1 undeclared; mean age = 31.49), 38 English (17 females, 17 males, 4 undeclared; mean age = 41.00), and 42 Spanish writers (25 females, 16 males, 1 undeclared; mean age = 32.45). The Chinese participants were all born and raised in Hong Kong, Taiwan, or Mainland China and reported Chinese as their native language; the English participants were all born and raised in the U.S. and reported English as their native language; and the Spanish participants were all born and raised in Central or South America, and reported Spanish as their native language. All reported being fluent in reading, writing, and speaking in their target language.

Participants were recruited from the local communities in the San Francisco Bay Area through online and hardcopy ads seeking individuals who were 18 years of age or older, and born and raised in a country for which the target language was the official language. The ads stipulated that the study requires reading and writing in the target language, and that all participants must read and write the target language fluently. Ads appeared in both English and in the target languages. Prior to participation all potential participants were telephone screened according to the inclusion criteria recruited for, and answered the same questions in a standard demographics form obtained as part of the pre-session measures, including self-reported reading, writing, and speaking proficiencies (see below).

Age differed significantly among the three groups, $F(2, 118) = 7.10$, $p = .001$, $\eta_p^2 = .11$. To examine if age covaried with the dependent variables tested, we computed correlations between age and the coded statement analysis categories (described below), separately for the true and false statements and each of the groups. Of the 66 effects (11 coded categories x 2 statements x 3 language groups), only 3 were significant (2 for the Chinese, 1 for English, and 0 for Spanish). Thus we concluded that age differences did not covary with the differences in usage of the statement analysis categories.

Measures

At the beginning of the experiment all participants completed the following instruments:

- A basic demographics questionnaire that confirmed ethnic group identity, places of birth and upbringing of themselves and parents, and first and other languages with self-ratings of language proficiency (excellent, good, fair, poor separately for reading, writing, and speaking)
- The General Ethnicity Questionnaire (GEQ; Tsai, Ying, & Lee, 2000); see below.
- An emotion checklist (guilt, fear, anger, embarrassment, worry, contempt, excitement, disgust, amusement, nervousness, surprise, and interest) in which participants self-reported their emotional experiences using 9-point scales labeled 0, None, 4, Moderate Amount, and 8, Extremely Strong. Participants also completed this checklist at the end of the experiment.
- The Machiavellianism Scale (Christie, 1970), a 10-item test assessing individual differences in cunning, duplicity, or interpersonal manipulation. Previous studies have demonstrated its internal reliability and convergent and discriminant validity with other personality measures (Paulhus & Williams, 2002).
- The NEO-Five Factor Inventory (Costa & McCrae, 1992), a 60-item test assessing the five personality traits found to be universal (Costa & McCrae, 1992): Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. Participants respond to each item using a 5-point scale (0 = strongly disagree to 4 = strongly agree); scale scores are computed using a standard formula. There is substantial evidence for the cross-cultural equivalence in the factor structure and within-country validity of the NEO-FFI (McCrae & Costa, 1997; McCrae, et al., 2005).
- The Social Dominance Orientation Scale (Pratto, Sidanius, Stallworth, & Malle, 1994), a 16-item test that measures individual differences in preferences for hierarchies within social groups and dominance of lower-status groups. After reverse coding specific items, all items are summed to produce a score. There is ample evidence for the internal and temporal reliability, and predictive and discriminant validity of the scale (Pratto, et al., 1994).
- The Self-Monitoring Scale (Snyder, 1974), a 25-item, true-false scale that assesses individual differences in expressive self-presentation and impression management. There is ample evidence for the internal and temporal reliability of the scale, along with its predictive validity (Lennox & Wolfe, 1984; Snyder, 1974).

The GEQ is a commonly used scale to measure acculturation and ethnic identity, and was included as a manipulation check for ethnic/cultural differences. It contains 38 statements, 25 rated on a 5-point Likert scale from strongly disagree to strongly agree and 13 rated on a 5-point scale from very much to not at all. The target group mentioned in the GEQ was modified to be applicable to each ethnic group. Analyses of the GEQ Total score, which was the mean of all items after reverse coding those negatively loaded, indicated that the Chinese sample had significantly higher scores than American born Chinese reported by Tsai, et al. (2000), $t(42) = 3.04$, $p = .004$, $d = .46$, demonstrating that our Chinese sample identified themselves as Chinese and strongly with Chinese culture

more so than American born Chinese. GEQ norms for Hispanics do not exist but their scores were comparable to the Chinese in our sample.

Stakes

Many studies in the deception literature have examined lies produced in situations in which participants were not very motivated one way or another to lie or tell the truth because they did not believe there were rewards or punishments associated with their performances. Higher-stakes studies are more analogous to real-life situations that face law enforcement and security personnel, and the behavioral indicators associated with veracity and lying that emerge from higher-stakes studies are different and more compelling than those from lower-stakes studies (DePaulo, et al., 2003; Frank & Svetieva, 2013). Identifying indicators that are based in low-stakes studies that are not analogous to real-life situations, or that are otherwise not validated, and then training law enforcement personnel on them would have dire consequences. At least one study has demonstrated detrimental effects of training to detect lies when non-validated indicators are used (Kassin & Fong, 1999). Consequently in this study participants were led to believe that their statements would be read by investigators who would make a determination about the believability of the statements, and that there were rewards and punishments for the participants depending on those determinations.

The stakes associated with their performances were as follows:

- If they lie about what they witnessed and wrote and are believed, they will receive an additional \$75 and will be allowed to leave early.
- If they lie about what they witnessed and wrote but are not believed, they will receive no additional money and will have to stay an additional hour filling out a long questionnaire.
- If they tell the truth about what they witnessed and wrote and are believed, they will receive an additional \$10 and will be allowed to leave early.
- If they tell the truth about what they witnessed and wrote but are not believed, they will receive no additional money and will have to stay an additional hour filling out a long questionnaire.

The stakes were different for lying and telling the truth because they reflected the stakes that occur in real life for the type of investigative interview examined in this study. Being a successful liar is likely associated with relatively large rewards in real life; and participants who may not be inclined to do so in the first place require additional motivation to do so. As it is easier for people to tell the truth, there are indeed rewards for telling the truth, but they are lower than when successfully lying. If, however, the type of investigative interview were different, different stakes might be more appropriate. For example, if participants were falsely accused of lying about what they had witnessed, there would be a larger stake in being perceived as truthful. This, however, would require a different study. Thus readers are cautioned to interpret the findings reported below vis-à-vis the particular way in which the experiment was conducted, including the stakes involved.

Procedures

Upon arrival to the laboratory, participants were led into a private instruction and consent area with a computer workstation. All forms, protocols, questionnaires, surveys, and instructions were presented in the native language of the participant (i.e., English, Spanish, or Chinese). After informed consent was obtained, participants were left alone to complete the pre-session measures via web-based surveys on the computer. When completed, participants rang a bell to call the experimenter back into the room. The instructions were then presented to the participants via audio PowerPoint.

The instructions informed participants that they will view a video that may or may not portray a crime, and then be asked to write two statements about what they witnessed. Participants were not specifically told at this time that they would be writing both a false and a true statement and in what order. Separate pilot testing of the procedures indicated an order effect when participants are asked to produce true and false statements. Participants in that pilot study who wrote a true statement first and then a false one reported that the content of their false statements was heavily influenced by their knowledge of what they had just written in the true statement. Those who wrote a false statement first, however, reported that their false statements were uncontaminated by their knowledge of what they may have written in a true statement (because they wrote the false statement first). True statements written after a false statement were not affected by order. Moreover, uncontaminated false statements written first are more likely to have ecological validity. Thus we opted to have all participants write both statements in a fixed false-true order, which was at first unbeknownst to the participants.

Participants were told that one of their statements will be selected to be read and evaluated by experts, that they may also be interviewed about what they witnessed and wrote, and that a determination will be made about whether they are lying or telling the truth. They were informed about the stakes associated with their performance, as described above. They were told that there are big rewards if they are believed, but also serious consequences if they are not believed; thus they must try to be as convincing as possible. After the instructions were delivered, including the stakes, and any questions answered, participants rated the severity of the consequences they faced in the experiment using a 10-point scale. Mean ratings across all participants was 5.13, commensurate with the participants' perception of moderate level of stakes in the experiment.

After ensuring that participants understood the instructions and stakes, the first experimenter left and a second experimenter entered and escorted participants to a second room in which there was a computer workstation. The experimenter asked the participants to follow the on-screen instructions, which read as follows:

"You will now see a video that may or may not show a crime. There is no audio in the video. You should play the video only once and you cannot take notes. You can click the bottom right corner of the image to view the video in full screen. Click Next when you're ready to watch the video."

Video 4 then played in its entirety, lasting 49 s. When the video finished, participants were then instructed to write a statement about what they saw in the video, as follows:

“Soon you will be interviewed by a security officer about what you witnessed. Before that interview, you first need to write a FALSE statement about what you witnessed in the video. Please write this FALSE statement about what you have just witnessed knowing that the statement may be read by a security officer and others who will determine whether it is believable or not. You may also be interviewed based on your statement. Thus write this FALSE statement to be as believable as possible. Use the paper and pen provided. You can write as much or as little as you want.

Write this statement in your native language. Do NOT go back and review the video.

When you are done, please ring the bell and wait for the experimenter. Do NOT click Next.”

Lined paper and a pen were provided. The participant wrote the statement, and then rang a bell when done to call the experimenter back into the room. The experimenter ascertained that the participants understood the instructions (i.e., wrote a FALSE statement), and then took the statement and labeled it “A” in plain sight of the participants. The experimenter then instructed the participants to click to the next screen after the experimenter left the room. The following instructions appeared:

“Now please write a TRUE statement about what you witnessed in the video. Remember, this TRUE statement may be read by a security officer and others who will determine whether it is believable or not. You may also be interviewed based on your statement. Thus, write this TRUE statement to be as believable as possible. Use the paper and pen provided. You can write as much or as little as you want.

Write this statement in your native language. Do NOT go back and review the video.

When you are done please ring the bell and wait for the experimenter. Do NOT click Next.”

When participants rang the bell, the experimenter re-entered the room, ascertained that participants understood the instructions (i.e., wrote a statement a TRUE statement), and then took the statement and labeled it “B” in plain sight of the participants.

Immediately after the writing exercise was completed, one of the statements was indeed selected and participants were interviewed about their statements using a standard interview protocol (i.e., questions were designed a priori and administered uniformly to all participants). As the analysis of the interview is not part of this study no further mention of it will be made.

After completion of the interview the second experimenter escorted the participants back to the initial instructions and consent area with the first computer workstation. The experimenter exited and the first experimenter re-entered. Participants were asked to complete some brief post-session measures on screen, were debriefed, compensated, and excused.

Coding

Structural Balance. The number of lines devoted to the prologue, incident, and epilogue portions of each statement was counted, and statements were coded dichotomously as either balanced or unbalanced. Balanced statements were defined to contain at least 33% of the total lines in the statement, with at least 20% devoted to each of the prologue and epilogue.

Word Count. The total word count for each statement was tallied.

Unique Sensory Detail (USD) and Spatial Detail (SD). The number of sentences in each statement that contained evidence for either/or both USD and SD – that is, specific descriptions generated by the five sensory perceptions to include sight, sound, touch, smell, taste and touch, or specific locations and the physical relationships of objects, people, etc., in relation to one another – was counted.

Emotion. The number of sentences within the epilogue of the statement that contained words that described the writer's emotion was counted.

Extraneous Information. Each sentence within a participant's response that contained extraneous information was identified, regardless of the extent of the extraneous information within that one sentence, and the total number of sentences within each statement was tallied.

Equivocation. The number of words or phrases within each statement that were construed as equivocation words/phrases from the writer's vantage point was counted. Equivocation pertained to actions by individuals in the video were not counted. For example, the statement "*The motorcyclist was sort of responsible for what happened*" was not counted as equivocation because the equivocation pertained to an individual in the video and not to the writer's perception of what happened. But the statement "*I sort of recall the vehicle being white*" was counted as equivocation because it pertained to the writer's perception.

Non-Prompted Negation (NPN). The number of words or phrases within each statement that were construed as NPN as they pertained to the writer was counted. For example, the sentence "*The motorcycle did not cross the road*" was not counted as NPN because it pertained to an action by an individual in the video and not the writer. The statement "*I did not see a motorcycle*" was counted as NPN because it pertained to the writer's actions or perceptions.

Moderating adverbs. Each word that constituted an Editing, Minimizing, or Intensifying adverb within a response was identified, and the total number of instances within each statement was tallied for each of these three types of adverbs. Adverbs that were counted had to pertain to the actions or perceptions of the writer; adverbs that pertained to activity by the individuals in the video were not counted.

Passive voice. The number of uses of the passive voice within each statement was counted.

Coding procedures and reliability. Two coders coded the linguistic features of the statements. One coder (VAS) had several decades of law enforcement experience and extensive experience in conducting statement analysis in real-life investigative settings,

was fluent in English and Spanish and coded the English and Spanish statements. A second coder, also an individual with several decades of experience in a law enforcement agency, was fluent in English and Chinese and coded the English and Chinese statements. Both coders first independently coded statements from 20 randomly selected English statements (10 true and 10 false). Initial reliabilities (Kappa for Structural Balance, ICCs for all other categories) were calculated on the initial set of 20 statements and ranged from .74 to 1.00. The coders were then instructed to arbitrate any disagreements and recalibrate their codes. They then independently coded the transcripts and statements from a new set of 20 English statements. Reliabilities computed across all 40 statements coded were high and acceptable for all coding categories ($.79 < \text{interrater reliability} < 1.00$). Both coders then completed coding the remaining English statements, and then coded either the Spanish or Chinese statements. Statements were provided with no marks or indicators of condition.

When the writer made a very obvious typographical error and it was readily apparent from the context what the writer intended (e.g., “cor” instead of “car”), the misspelled word was analyzed and included in the word count (WC). If a determination about what the writer meant in the use of the misspelled word could not be made from the context, the word was still treated as a word for word count purposes but was not marked as any other applicable linguistic feature. Also, when writers crossed out words, phrases, or sentences and they could clearly be deciphered, they were analyzed for linguistic features and were included in the word count. If a determination about what the writer meant in the use of the crossed out words, phrases, or sentences could not be made, they were not marked as any applicable linguistic feature nor were they included in the word count.

Results

Main Analyses

We first computed aggregate scores for the veracity and deception indicators by summing the codes for Emotions and USD-SD (veracity indicators), and the codes for Extraneous Information, Equivocation, NPN, Moderating Adverbs, and Passive Voice (deception indicators), separately for each statement. We then computed a Language (3) by Veracity Condition (2) by Indicator Type (2) mixed ANOVA on the aggregate scores. The Veracity Condition by Indicator Type interaction was significant, $F(1, 120) = 14.50$, $p = .000$, $\eta_p^2 = .11$. As predicted true statements had more veracity indicators than did false statements, while false statements had more deception indicators than did true statements (Figure 1). Importantly, the Language by Veracity Condition by Indicator Type was *not* significant, $F(2, 120) = 2.30$, $p = .10$, $\eta_p^2 = .04$, indicating that language did not moderate the interaction between Veracity Condition and Indicator Type.

In order to analyze each of the coded linguistic features separately, we computed a mixed-factor MANOVA using Language and Veracity Condition as independent variables and the scalar coded linguistic features as multiple dependent variables. The main effects of Language and Veracity Condition were significant, $F(14, 226) = 7.68$, $p = .000$, $\lambda = .45$; and $F(7, 112) = 3.90$, $p = .001$, $\lambda = .80$, respectively; the interaction was not, $F(14, 224) = 1.43$, $p = .15$, $\lambda = .84$. Follow up, univariate analyses decomposing the Veracity Condition effect indicated that true statements contained more elements of

USD-SD than did false statements, while false statements contained more minimizing adverbs than did true statements (both $ps < .05$; Table 2). For Structural Balance, we computed one-sample binomial tests on the proportion of statements that were coded balanced or unbalanced, separately for the true and false statements. As predicted a greater proportion of the false statements (77%) were coded as unbalanced, $z = 5.95$, $p = .000$; for true statements there were no differences between the proportion of statements coded as balanced (55%) or unbalanced, $z = 1.09$, $p = .28$.

Post Hoc Analyses

To decompose the significant main effect of Language, we computed follow-up, oneway analyses with pairwise comparisons using Bonferroni corrections. Chinese writers used more words, sensory and spatial detail, and intensifying adverbs than did the English and/or Spanish samples, while the English writers used more words conveying extraneous information (Table 3).

General Discussion

As a whole the selected linguistic features discriminated between true and false witness statements and the effect sizes were relatively large. Univariate analyses indicated that structural balance, unique sensory and spatial detail, and minimizing adverbs were particularly important in differentiating true and false statements. Importantly, language did *not* moderate the relationship between veracity and the coded features.

The findings were not produced without limitations, the first of which concerned the laboratory setting in which the crimes were viewed. Actually witnessing a hit and run occur, with the associated sensory perceptions (e.g., sights, sounds, smells, etc.) would have been more realistic. Also, the task asked of the participants (i.e., to write a false statement) may have been too general. If individuals actually had to lie about what they witnessed, they would do so for a specific reason, such as wanting to conceal the identity of the driver of the car that did the hit and run, or needing to deny being in that particular place at that particular time. We purposely did not give these kinds of instructions to the participants because if we did, truth tellers would also have to lie because they would have had to put themselves into the hypothetical situation that we instructed them. This would have resulted in the true statements not really being true. Thus we gave participants the flexibility about exactly how to craft their false statement. The limitation in doing so, however, is that some participants only changed a seemingly minor or irrelevant detail in the false statement, which was likely not the type of statements one would obtain if one had a very specific reason for lying about having witnessed a crime, and which may not require the kinds of linguistic or grammatical choices when producing more realistic false statements. These limitations probably led to a dilution of the quality of the statements we obtained and analyzed. On one hand, this may have influenced the degree to which many of the linguistic features were observable or not in the statements, which in turn resulted in many near-zero categories of data (which was indeed observed). On the other hand, the limitations may have also made it more difficult to obtain statistical significance, which would be an acceptable Type II error. The fact that we obtained positive findings even with such statements obtained in such contexts may have been even more meaningful.

Another limitation concerned the specific features selected for coding in this study. As mentioned previously, we selected only a few SA categories that were deemed appropriate in this study given the experimental context and procedures, and for which there was previous empirical support. Different coded features may lead to the different outcomes.

A final limitation may have occurred because of the fixed order of the tasks in the study. It may have been possible, for example, for participants to provide inaccurate information in their true statements because of a bias of having written a false statement prior to writing the true statement. To mitigate against this possibility, we reviewed all the statements provided by the participants. We confirmed that all true statements were indeed accurate depictions of what occurred in the crime video they witnessed (albeit with large individual differences in the amount of details reported). We also confirmed that all false statements did indeed contain some kind of false information. Thus this bias for the potential recall of inaccurate information in the true statements did not occur in our study.

Regardless of these limitations the findings provided exciting initial evidence for the potential cross-cultural, cross-language generalizability of SA to differentiate truths from lies. That the SA features did indeed differentiate true and false statements was not new to the literature; the unique contribution of this study was the fact that multiple languages were tested in the same study and that language did not moderate the ability of the SA features to differentiate truths from lies. Although a number of studies have examined the ability of linguistic features to do so in different languages separately (Masip, et al., 2012; Ruby & Brigham, 1997; Schelleman-Offermans & Merckelbach, 2010), this study is the first to examine multiple languages in the same experiment, providing for an apples-to-apples comparison across languages and findings.

As mentioned earlier, cross-language consistency in the relationship between SA features and veracity suggests a potential pancultural similarity in the underlying psychological effects of lying, and similarity in the linguistic choices that mark those effects. Although it is not known whether memory is structured similarly across cultures, the current findings suggest the existence of a possible universal mechanism underlying the psychological demands placed on individuals when lying, and in the linguistic and grammatical choices that individuals make when lying. If the rules of grammar and deep structure of language (Chomsky, 1957, 1972) and the principles of memory and recall (Undeutsch, 1989) are similar across cultures, and if there is cross-cultural similarity in the psychological demands placed on individuals when lying, then verbal indicators of truths and lies may occur regardless of culture, ethnicity, and language.

Cross-language similarity in the applicability of SA features has important potential practical implications. For example, the U.S., as many countries of the world, is a multi-ethnic, multi-language society, and a popular destination of peoples from many other countries of the world. Knowing that the principles and techniques of SA are applicable across cultures and languages would be a major boon to those individuals whose jobs rely on making determinations about truths and lies – such as law enforcement officers, customs and immigration officials, or airport security personnel – who interact with people speaking different languages on a daily basis.

It is interesting to speculate about why some of the coded SA categories did not differentiate true and false statements. It is very likely that some of the limitations discussed above concerning the nature of the task and the setting within which the crime was viewed limited the nature of the statements produced. Making the task more realistic and personal for the writers in the future may help to address this issue and allow for a fairer test of the various SA categories to differentiate truths from lies.

Although language did not moderate the relationship between SA features and veracity, there were some interesting language main effects. These effects have also been reported in a study of SA features coded in interviews and written statements in an experiment using a mock crime scenario (Matsumoto, et al., 2013). Although these findings clearly need to be replicated, they suggested cultural differences in the use of language that facilitate the use of some grammatical features but not others. These differences likely contribute to differences in overall communication styles.

Future research will need to replicate the findings obtained in this study in different contexts, with different languages, and different sources of statements (e.g., oral vs. written). In particular examining cross-linguistic differences in truths and lies related to one's own actions, not as a witness to someone else's actions, in a context in which participants are motivated and there are stakes for performance, may be a key test of the replicability of the findings reported here. Examining individual differences in the use of language across cultures in deceptive situations is also warranted.

References

- Adams, S. H. (1996). Statement analysis: What do suspects' words really reveal? *FBI Law Enforcement Bulletin*, 65, 12-20.
- Adams, S. H., & Jarvis, J. P. (2006). Indicators of veracity and deception: An analysis of written statements made to police. *Speech, Language, and the Law*, 13, 1-22.
- Chomsky, N. (1957). *Syntactic structures*. New York: Walter de Gruyter.
- Chomsky, N. (1972). *Language and mind*. New York: Harcourt, Brace, and Johanovich.
- Christie, R. (1970). Scale construction. In R. Christie & F. L. Geis (Eds.), *Studies in Machiavellianism* (pp. 10-34). New York: Academic Press.
- Connelly, S., Allen, M. T., Ruark, G. A., Kligyte, V., Waples, E. P., Leritz, L. E., & Mumford, M. D. (2006). Exploring content coding procedures for assessing truth and deception in verbal statements, Year 3, Cumulative Final Report. Norman, OK: University of Oklahoma.
- Costa, P. T., & McCrae, R. R. (1992). *Revised Neo-Personality Inventory (NEO-PI-R) and Neo Five Factor Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129, 74-118.
- Frank, M. G., & Svetieva, E. (2013). Deception. In D. Matsumoto, M. G. Frank & H. S. Hwang (Eds.), *Nonverbal Communication: Science and Applications* (pp. 121-144). Newbury Park, CA: Sage.
- Hauch, V., Blandon-Gitlin, I., Masip, J., & Sporer, S. L. (2012, 23-27 April 2012). *Linguistic cues to deception assessed by computer programs: A meta-analysis*. Paper presented at the Proceedings of the EACL 2012 Workshop on Computational Approaches to Deception Detection, Avignon, France.
- Johnson, M. K. (1988). Reality monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General*, 117, 390-394.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.
- Johnson, M. K., & Raye, C. L. (1998). False memories and confabulation. *Trends in Cognitive Sciences*, 2, 137-145.
- Kassin, S. M., & Fong, C. T. (1999). "I'm innocent!": Effects of training on judgments of truth and deception in the interrogation room. *Law and Human Behavior*, 23, 499-516.
- Lennox, R. D., & Wolfe, R. N. (1984). Revision of the Self-Monitoring Scale. *Journal of Personality and Social Psychology*, 46, 1349-1364.
- Masip, J., Bethencourt, M., Lucas, G., Sanchez-San Segundo, M., & Herrero, C. (2012). Deception detection from written accounts. *Scandinavian Journal of Psychology*, 53, 103-111.

- Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime, and Law*, 11, 99-122.
- Matsumoto, D., Hwang, H. C., & Sandoval, V. A. (2013). Ethnic similarities and differences in linguistic indicators of veracity and lying in a moderately high stakes scenario. *Journal of Police and Criminal Psychology*. doi: 10.1007/s11896-013-9137-7
- McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist*, 52, 509-516.
- McCrae, R. R., Terracciano, A., Khoury, B., Nansubuga, F., Knezevic, G., & Djuric Jovic, D. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*, 88, 547-561.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29, 665-675.
- Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, 36, 556-563.
- Porter, S., Birt, A. R., Yuille, J. C., & Lehman, D. R. (2000). The negotiation of false memories: The influence of interviewer and rememberer characteristics on memory distortion. *Psychological Science*, 11, 513-516.
- Porter, S., & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior*, 20, 443-459.
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social Dominance Orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, 67, 741-763.
- Rabon, D. (1994). *Investigative Discourse Analysis*. Durham, NC: Carolina Academic Press.
- Ruby, C. L., & Brigham, J. C. (1997). The usefulness of the criteria-based content analysis technique in distinguishing between truthful and fabricated allegations: A critical review. *Psychology, Public Policy, and Law*, 3, 705.
- Rudacille, W. C. (1994). *Identifying lies in disguise*. Dubuque, IA: Kendall-Hunt.
- Sapir, A. (1996). *The laboratory for scientific interrogation course on scientific content analysis (SCAN) workbook*. Phoenix, AZ: Laboratory for Scientific Interrogation.
- Schafer, J. R. (2007). *Grammatical differences between truthful and deceptive written narratives*. Unpublished doctoral dissertation.
- Schelleman-Offermans, K., & Merckelbach, H. (2010). Fantasy proneness as a confounder of verbal lie detection tools. *Journal of Investigative Psychology and Offender Profiling*, 7, 247-260.

- Snyder, M. (1974). Self-Monitoring of Expressive Behavior. *Journal of Personality and Social Psychology*, 30, 526-537.
- Sporer, S. L. (2004). Reality monitoring and detection of deception. In P. A. Granhag & L. A. Stromwall (Eds.), *The detection of deception in forensic contexts* (pp. 64-80). New York: Cambridge University Press.
- Tsai, J. L., Ying, Y.-W., & Lee, P. A. (2000). The meaning of "being Chinese" and "being American": Variation among Chinese-American young adults. *Journal of Cross-Cultural Psychology*, 31, 302-332.
- Undeutsch, U. (1989). The development of statement reality analysis. In J. C. Yuille (Ed.), *Credibility Assessment* (pp. 101-119). New York: Kluwer Academic/Plenum Publishers.
- Vrij, A. (2007). Criteria based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, 11, 3-41.
- Vrij, A., & Mann, S. (2006). Criteria-based content analysis: An empirical test of its underlying process. *Psychology, Crime, and Law*, 12, 337-349.
- Weiner, M., & Mehrabian, A. (1968). *Language within language: Immediacy, a channel in nonverbal communication*. New York: Appleton-Century-Crofts.
- Weintraub, W. (1989). *Verbal behavior in everyday life*. New York: Springer.
- Zaparniuk, J., Yuille, J. C., & Taylor, S. (1995). Assessing the credibility of true and false statements. *International Journal of Law and Psychiatry*, 18, 343-352.

Table 1

ICCs across the 14 Emotion Means using Countries as Raters

<u>Video</u>	<u>ICC for absolute agreement</u>	<u>ICC for consistency</u>
1	0.890	0.930
2	0.902	0.937
3	0.910	0.941
4	0.913	0.948
5	0.906	0.945
6	0.910	0.944
7	0.927	0.957
All	0.913	0.946

Table 2

Means (top entry), Standard Deviations (bottom entry), and Cohen's d for each of the Coded Linguistic Features, Separately for True and False Statements

<u>Coded Feature</u>	<u>True Statements</u>	<u>False Statements</u>	<u>Cohen's d</u>
Word Count	97.05 (62.59)	92.12 (74.05)	-.10
USD-SD	2.32 (1.61)	1.81 (1.51)	.46
Emotion	.11 (.31)	.10 (.44)	.02
Extraneous Information	.08 (.31)	.07 (.25)	.06
Equivocation	.14 (.54)	.17 (.55)	-.06
Non-Prompted Negation	.06 (.30)	.07 (.43)	-.03
Editing Adverbs	.02 (.13)	.02 (.16)	-.05
Minimizing Adverbs	.00 (.00)	.04 (.24)	-.17
Intensifying Adverbs	.62 (1.29)	.69 (1.45)	-.06
Passive Voice	.02 (.13)	.00 (.00)	.13

Note: USD-SD – Unique Sensory Detail and Spatial Detail

Table 3

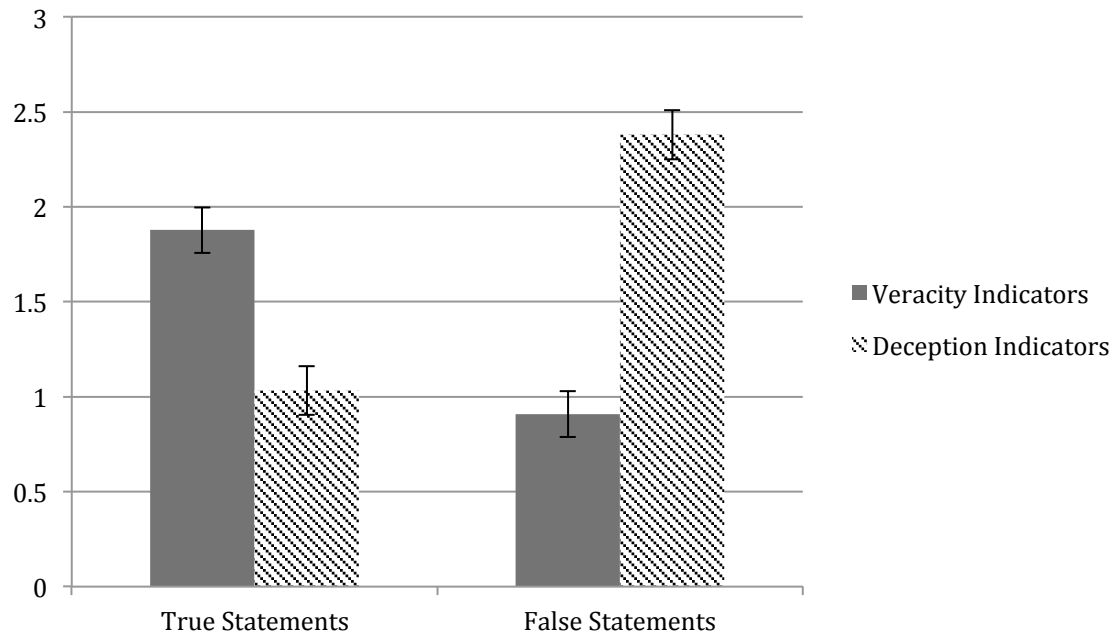
Means (top entry), Standard Deviations (bottom entry), and Results of Pairwise Tests of Language Differences with Bonferroni Corrections

<u>Variable</u>	<u>Chinese</u>	<u>English</u>	<u>Spanish</u>	<u>Bonferroni Results</u>
Word Count	233.30 (154.13)	162.45 (110.76)	160.40 (101.32)	Chinese > English
USD-SD	5.49 (2.81)	4.05 (2.92)	2.62 (2.36)	Chinese > Spanish
Emotions	.33 (.78)	.24 (.49)	.05 (.31)	
Extraneous Info	.00 (.00)	.45 (.79)	.02 (.15)	English > Chinese, Spanish
Equivocation	.42 (.98)	.45 (1.43)	.05 (.22)	
Non-Prompted Negation	.14 (.64)	.21 (.58)	.05 (.22)	
Editing Adverbs	.09 (.37)	.03 (.16)	.00 (.00)	
Minimizing Adverbs	.07 (.34)	.05 (.23)	.00 (.00)	
Intensifying Adverbs	3.12 (3.41)	.63 (1.20)	.00 (.00)	Chinese > English, Spanish
Passive Voice	.00 (.00)	.05 (.23)	.00 (.00)	

Note: USD-SD – Unique Sensory Detail and Spatial Detail

Figure 1

Differences between True and False Statements as Measured by Veracity and Deception Indicators (error bars refer to Standard Errors)



Appendix 3 - Report on Main Study, Year 3 (currently submitted to a peer-reviewed journal)

Running Head: Lies in Linguistic Cues

Linguistic Cues of Deception across Multiple Language Groups

Hyisung C. Hwang

David Matsumoto

San Francisco State University and Humintell

Keywords: Language, Statement Analysis, Ethnicity, Deception Detection, Crime

CORRESPONDING AUTHOR: Address all correspondence concerning this manuscript to

Hyisung C. Hwang, Ph.D.

Department of Psychology

San Francisco State University

1600 Holloway Avenue

San Francisco, CA 94132

Telephone: +1-415-338-1114

Fax: +1-510-217-9608

Email: hwang@sfsu.edu

AUTHOR BIOS:

Hyisung C. Hwang is an adjunct faculty at San Francisco State University and a research scientist at Humintell. Her research interests are in emotion, deception detection, statement analysis, nonverbal behaviors, and culture/ethnicity.

David Matsumoto is Professor of Psychology and Director of the Culture and Emotion Research Laboratory at San Francisco State University. He has studied culture, emotion, nonverbal behavior, deception detection, social interaction and communication for over 30 years.

DECLARATION OF CONFLICT OF INTEREST:

The author(s) declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

FINANCIAL DISCLOSURE/FUNDING:

This report was prepared with the support of research grant FA9550-11-1-0306 from the U.S. Air Force Office of Scientific Research.

ACKNOWLEDGMENTS OR AUTHOR NOTES:

The authors wish to thank Vincent A. Sandoval and Ling-Ling Beerman for their assistance as coders; and Jenna Propst, Laurel Richards, Sophia Nguyen, Tram Nguyen, and Isaac Perry, for their assistance in collecting data used in this study.

Abstract

A recent study (Matsumoto & Hwang, in press) showed that specific linguistic and grammatical features of a technique commonly referred to as Statement Analysis are applicable across different language groups. One limitation of that study was that it used an eyewitness crime video paradigm, which might be different from writing a statement after committing an actual criminal act. We remedied that limitation in this study by using a mock crime paradigm. In this study, three language groups (English, Spanish, Chinese) produced statements after committing a mock crime, taking a check, in an experimental context. The results indicated that certain linguistic features significantly functioned as indicators discriminating truths from lies across different language groups, suggesting that SA might be applicable as a reliable indicator of deception across languages.

Key words: Language, Statement Analysis, Ethnicity, Deception Detection, Crime

Linguistic Cues of Deception across Multiple Language Groups

Understanding verbal deception cues from written statements is one way to distinguish truths from lies. This method is critically important in investigative contexts because written statements are commonly required and expected, and the available statements from witnesses and suspects can be used to increase the efficacy of the interview. Thus the ability to analyze the credibility of statements provided is useful for interrogators and interviewers who are interested in distinguishing truthful information from false ones.

Although analyzing written statements is practically useful in detecting deception (Harlow, 2014), the ability to make accurate decisions about the veracity or deception of statements is complicated because writing allows people to have time to think and sort their thoughts and feelings, which reduces the risks of leaking errors and obvious flaws in lying. So there may be less noticeable leakage of indicators of deception if writers are extremely prepared and experienced. Despite the tricky challenge of distinguishing truthful statements from lying ones, doing so is still important and necessary as written statements can often be the most promising gate keeper in conducting investigative interviews.

One class of techniques for analyzing statements for veracity and deception is known as Statement Analysis (SA; aka scientific content analysis, investigative discourse analysis; Leo, 2008). SA is defined an effective technique to guide investigative interviews (Vrij, 2008), and is a broad concept that includes specific systems such as Criteria Based Content Analysis (CBCA), Psychological Narrative Analysis (PNA), and Reality Monitoring (RM). SA had its roots in psycholinguistic research in the early 1900s and its more modern roots in the work of Undeutsch (1989) and a technique known as Statement Validity Analysis (SVA), which was based on a hypothesis that statements associated with actual memories differ from statements based on fabrication or fantasy (Undeutsch, 1989).

Many studies have been conducted on identifying linguistic cues of deception via SA (Ruby & Brigham, 1997; Porter & Yuille, 1996; Sporer & Schwandt, 2006; Duran, Hall, McCarthy, & Mcnamara, 2010; Masip, Bethencourt, Lucas, Segundo, & Herrero, 2012). CBCA is the one of the most studied strategies; it has 19 criteria such as general, unusual, motivational and stylistic features (Undeutsch, 1954), which can be flexible depending on usage. A person who scores higher numbers of the criteria indicates the probability that he/she is being honest in statements (Colwell, Hiscock-Anisman, & Fede, 2013). Willén and Strömwall (2012) found that some individual CBCA criteria indeed differentiated truths from lies, but the total scores of CBCA were not able to distinguish truths from lies in their study. Johnson and Raye (1981) believed that more external-sensorial information and contextual information would appear in memories of actually experienced events. RM-based techniques have led to accurate rates in the 80% range when predicting statements as honest or deceptive (Masip et al., 2005).

As briefly reviewed above, much research and valuable findings have been produced regarding statement analysis, which have contributed to substantial empirical evidence of its validity. However, SA has been criticized because of inadequate evidence as to its application to various languages as most empirical evidence was derived from

the original language (e.g., German, English) in which it was developed (Leo, 2008). Considering the efficacy of SA, there is a great possibility and need of its utility in various languages. Extending the usability of SA to various languages is demanding but meaningful as it is one way to examine the reliability of SA.

In order to address this gap in the literature, a few studies have examined indicators of deception in languages other than English (Masip, Bethencourt, Lucas, Sanchez-San Segundo, & Herrero, 2012; Ruby & Brigham, 1997; Schelleman-Offermans & Merckelbach, 2010; Spence, Villar, Gina, & Arciuli, 2012). One limitation of these previous studies, however, was that each studied a different language and no one study compared different languages within the same study using the same methodology. Thus, although these earlier studies were suggestive of the potential cross-language applicability of SA, comparing results across them is problematic because study differences confound the languages examined.

A more recent study addressed this limitation (Matsumoto & Hwang, in press). In that study, participants from three language groups (English, Spanish, Chinese) witnessed a video portraying an actual crime and then wrote false and true statements about what they had witnessed in their respective languages. Selected linguistic features of SA discriminated between true and false witness statements, and language did not moderate the relationship between veracity and the coded features.

This latest study described above contributed to the scientific evidence by showing that specific and reliable linguistic and grammatical features of a SA technique were applicable across multiple language groups. However, the study was also limited because writing about having witnessed a crime video may differ from actually experiencing and committing a criminal act. To extend that study and to remedy this particular limitation, we conducted a study by using a mock crime paradigm. In the study, participants from three language groups (English, Spanish, Chinese) produced statements after committing a mock crime, taking a check, in an experimental context. The Spanish and Chinese language groups were selected as they are the largest foreign language groups among immigrants and in the overall populations in the U.S. One reason for using a mock crime scenario is that the literature has emphasized the importance of stakes and motivation in lying in experimental contexts (DePaulo et al., 2003; Frank & Svetieva, 2013; Matsumoto & Hwang, in press). The written statements provided by the participants were analyzed using SA.

Linguistic Markers of Veracity and Lying used in this Study

In this study, the same SA categories that were tested in the previous study (Matsumoto & Hwang, in press) were used: unique sensory detail and spatial detail (USD-SD), extraneous information, equivocation, non-prompted negation, passive voice, moderating adverbs.

Indicators of veracity

Unique sensory detail and spatial detail (USD-SD). Unique sensory detail (USD) pertains to specific descriptions generated by the five sensory perceptions (sight, sound, touch, smell, taste and touch). Spatial detail (SD) pertains to specific locations and the physical relationships of objects, people, etc., in relation to one another (Adams & Jarvis,

2006). Truthful statements are expected to contain USD-SD details about a specific event. CBCA (Porter & Yuille, 1996; Undeutsch, 1989; Vrij, 2007) and the Reality Monitoring frameworks (Johnson, 1988; Johnson & Raye, 1981) have provided strong evidence to suggest that individuals who recall previously encoded events truthfully report more sensory and spatial details because these details are encoded in memory along with the factual content of the event.

Indicators of lying

Extraneous information. Extraneous information is information that does not answer the question posed, and may be used to justify the liars' actions, deflect the question because they may not want to respond to that specific question, help liars distance themselves from the act of lying or the content of the lie, or aid liars in exerting control over the interview (Adams, 1996). This idea has been supported by many studies (DePaulo et al., 2003; Matsumoto, Hwang, & Sandoval, 2013; Vrij, 2007)

Equivocation. Equivocation refers to information that is not relevant to the question that was posed to elicit the statement (e.g., maybe, kind of). Equivocation words qualify statements, allowing liars to distance themselves from the act or content of lying by tempering the action about to be described or by discounting the message even before it is transmitted (Weintraub, 1989). Matsumoto and colleagues (2013) reported that liars from different ethnic groups produced more equivocation when writing statements in English.

Non-prompted negation (NPN). Negation in discourse or statements may be an indicator of deception inasmuch as respondents may use it to carefully omit their involvement in a crime (Adams & Jarvis, 2006), and there are generally more negative statements in deceptive oral narratives than in truthful oral accounts (Hauch, Blandon-Gitlin, Masip, & Sporer, 2012; Newman et al., 2003; Porter et al., 2000). Matsumoto and colleagues (2013) reported that liars from different ethnic groups produced more NPN both when writing statements in English and in oral interviews.

Passive voice. When describing actions, people generally assume responsibility for those actions by employing the active voice. Passive voice occurs when the object of an action verb appears as the subject of the sentence. It may be used when liars attempt to conceal their identity as an actor, distancing themselves from the action of the verb (Connelly et al., 2006; Rudacille, 1994).

Moderating adverbs. Moderating adverbs consists of *Intensifying adverbs* (e.g., very, really, honestly), and are typically used when a communicator is attempting to convince another person of something, *Minimizing adverbs* (e.g., only, just) are used to minimize the role of the actor, *Editing adverbs* (e.g., after, next, so), indicating a temporal lacunae (Rabon, 1994; Schafer, 2007). Adverbs are often used to edit information that might be crucial to an inquiry. Matsumoto and colleagues (2013) reported that liars from different ethnic groups produced more moderating adverbs both when writing statements and in oral interviews.

Overview of the Study

Based on the recent findings demonstrating cross-language applicability of certain categories of SA as an indicator of veracity and deception (Matsumoto & Hwang, in

press), we hypothesized that the coded SA categories would differentiate truthful statements from lying ones across the three languages tested.

Methods

Participants

All participants were adults age 18 or older, and came from one of the three ethnic/language groups: European Americans, Chinese immigrants, and Hispanic immigrants. The European Americans were all born-and-raised in the U.S. and whose first language was English ($n = 35$ for males, $n = 28$ for females). The Hispanics were individuals who were born and raised in any country in Central or South America, or whose parents were born in any of those countries, and whose first language was Spanish ($n = 24$ for males, $n = 25$ for females). The Chinese participants were individuals born and raised in the People's Republic of China, Hong Kong or Taiwan, or whose parents were born and raised in those countries, and whose first language was Mandarin or Cantonese ($n = 16$ for males, $n = 40$ for females). As a manipulation check on language fluency, participants were asked to self-evaluate their reading and writing levels (poor to excellent) in the target language. Only participants who highly rated their reading and writing skills in the primary language were selected for the research participation. Additionally, participants' self-ratings of their ethnic group identity were checked by the measure of the General Ethnicity Questionnaire (GEQ; Tsai, Ying, & Lee, 2000). The statements from the participants who withdrew consent at the end of the experiment, or misunderstood their condition or experimental roles (e.g., forgot to enter the file room and take the check, did not write in their primary language, or confused with their assigned condition) were excluded.

Measures

At the beginning of the experiment, all participants completed a series of questionnaires including a brief demographic questionnaire, the General Ethnicity Questionnaire (GEQ), the Machiavellianism Scale (Christie, 1970) and the Self-Monitoring Scale (Snyder, 1974). Participants also completed an emotion checklist at the beginning and the end of the experiment. This checklist included 12 emotion words (guilt, fear, anger, embarrassment, worry, contempt, excitement, disgust, amusement, nervousness, surprise and interest) rated on nine-point scales labeled 0 = *None*, 4 = *Moderate Amount* and 8 = *Extremely Strong*.

The GEQ is a commonly used scale to measure acculturation and ethnic identity and was included as a manipulation check for ethnic/cultural differences. This questionnaire contains 38 statements, 25 rated on a five-point Likert scale from strongly disagree to strongly agree and 13 rated on a five-point scale from very much to not at all. The GEQ was modified to be applicable to each ethnic group. Analyses of the GEQ Total score, which was the mean of all items after reverse coding those negatively loaded, indicated that our Chinese sample had significantly higher scores than the American born Chinese and Chinese who immigrated to the US before the age of 12 reported by Tsai et al. (2000), $t(64) = 14.58, p < .001, d = .85$; $t(64) = 7.87, p < .001, d = .46$, respectively. These analyses demonstrated that our Chinese sample identified themselves as Chinese and very strongly with Chinese culture, more so than American born Chinese. Norms for Hispanics using this same measure do not exist, but their scores were comparable to the

Chinese in our sample.

Procedure

Pre-session. Upon arrival to the laboratory, participants were instructed about the study and completed the consent form. Participants completed the pre-session measures. They were then given detailed instructions on the experiment, which differed depending on their truth or lie condition. The truth condition required participants not to take a check made out to cash for \$200, and to tell the truth in the interviews and written statement. The lie condition required participants to take the check and lie in the interviews and written statement. The assignments were determined randomly prior to the participants' arrival to the laboratory. Participants were told that they would have interviews regarding what they did in the file room, where the \$200 check was located and they would have to persuade the interviewers about their honesty. Participants were told that they would earn a minimum of \$30 for their participation, and bonuses of anywhere from \$0 to \$50 depending upon their assigned condition and the judgments of the interviewers. In reality, all participants received a standard fee of \$40. After the introduction, participants completed the last two questions in the pre-session measures.

Interviews and statement. After the pre-session, participants were guided to move to an interview room for an initial screening interview. The initial interview was for the purpose of ascertaining participants' intent to commit a crime. Once the first interview was completed, participants waited nearby and then entered the file room, where the check was located. Depending on participants' veracity condition, they stole the check or left it where it was. After the file room, participants were escorted to the next interview. Prior to the interview, the interviewer asked participants to write a statement about what they did in the file room on lined papers. A pen and paper were provided and participants were given as much as they wanted. The interviewers left the interview room during the writing. Participants were instructed to write in their native language. Once participants finished their writing, they rang a bell and the interviewer re-entered the room and briefly reviewed it before starting the second investigative interview, asking standardized questions so as to investigate participants' veracity. The interviews were conducted in English. As the purpose of this study was to examine whether language moderated the ability of SA categories to differentiate true from false written statements, the interviews were not analyzed, and no further mention of them will be made.

Post-session. After completing the interview, participants were escorted to a debriefing room and completed the post-session measures. The aim of the experiment was explained and they were given the standardized compensation fee, \$40 and no punishment.

Coding

The SA categories described earlier were coded as follows:

Unique Sensory Detail (USD) and Spatial Detail (SD). The number of sentences in each statement that contained evidence for either/or both USD and SD – that is, specific descriptions generated by the five sensory perceptions to include sight, sound, touch, smell, taste and touch, or specific locations and the physical relationships of

objects, people, etc., in relation to one another – was counted.

Extraneous Information. Each sentence within a participant's response that contained extraneous information was identified, regardless of the extent of the extraneous information within that one sentence, and the total number of sentences within each statement was tallied.

Equivocation. The number of words or phrases within each statement that were construed as equivocation words/phrases from the writer's vantage point were counted.

Non-Prompted Negation (NPN). The number of words or phrases within each statement that were construed as NPN as they pertained to the writer were counted.

Passive voice. The number of uses of the passive voice within each statement was counted.

Moderating adverbs. Each word that constituted an Editing, Minimizing, or Intensifying adverb within a response was identified, and the total number of instances within each statement was tallied for each of these three types of adverbs. Adverbs that were counted had to pertain to the actions or perceptions of the writer; adverbs that pertained to activity by the individuals in the video were not counted.

Coding procedures and reliability. Statements were coded by two trained raters who were blind to the conditions of the participants in the experiment. One coder had several decades of law enforcement experience and extensive experience in conducting statement analysis in real-life investigative settings, was fluent in English and Spanish and coded the English and Spanish statements. A second coder, also an individual with several decades of experience in a law enforcement agency, was fluent in English and Chinese and coded the English and Chinese statements. Both coders first independently coded statements from 20 randomly selected English statements (10 true and 10 false). Initial reliabilities (Intra-Class Correlations- ICCs) were calculated on the initial set of 20 statements and ranged from .89 to 1.00. The coders were then instructed to arbitrate any disagreements and recalibrate their codes. They then independently coded the statements from a new set of 20 English statements. Reliabilities computed across all 40 statements coded were high and acceptable for all coding categories ($.87 < ICC < 1.00$). One coder then completed coding the remaining English statements, and then the Spanish statements; the other coder coded the Chinese statements. Statements were provided with no marks or indicators of condition.

When the writer made a very obvious typographical error and it was readily apparent from the context what the writer intended (e.g., "cor" instead of "car"), or if the writer crossed out words and the words were legible, the word was analyzed for linguistic features. If a determination about what the writer meant in the use of the crossed out words, phrases, or sentences could not be made, they were not marked as any applicable linguistic feature nor were they included in the word count.

Results

Main analyses

We computed descriptives of all SA variables (see Table 1) and aggregate scores for the veracity and deception indicators by summing the codes for Extraneous

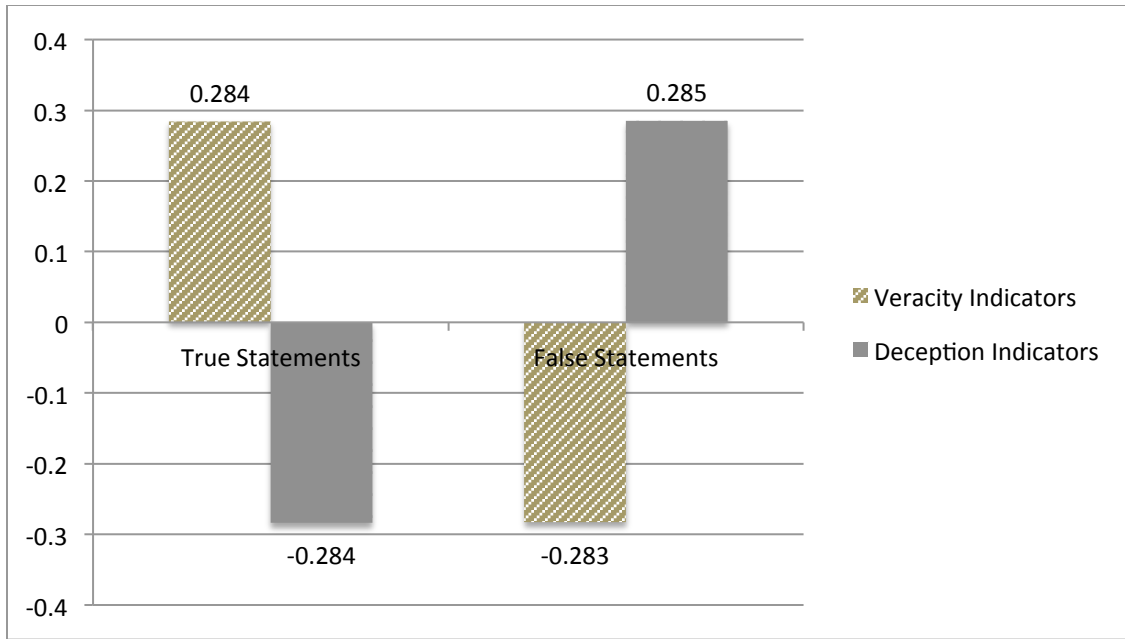
Information, Equivocation, NPN, Moderating Adverbs, and Passive Voice separately for each statement (USD-SD was used as the single veracity indicator). We then computed a Language (3) by Veracity Condition (2) by Indicator Type (2) mixed three-way ANOVA on the aggregate scores. The Veracity Condition by Indicator Type interaction was significant, $F(1, 157) = 9.827, p = .000, \eta_p^2 = .59$. As predicted true statements had relatively more veracity indicators than did false statements, while false statements had more deception indicators than did true statements (see Figure 1, which reports residualized means in order to present the pure interaction effect between Veracity Condition and Indicator Type; Rosnow & Rosenthal, 1989). Importantly, the Language by Veracity Condition by Indicator Type interaction was *not* significant, $F(2, 157) = .845, p = .431, \eta_p^2 = .011$, indicating that language did not moderate the interaction between Veracity Condition and Indicator Type.

Table 1. *Descriptives for all SA variables*

Variable	Truth	Lie
USD-SD	0.99	0.24
	(1.24)	(0.62)
Extraneous Information	0.36	.60
	(0.71)	(.88)
Equivocation	0.55	0.96
	(0.80)	(1.09)
Non-Prompted Negation	0.24	0.34
	(0.59)	(0.55)
Passive Voice	0.00	0.04
	(0.00)	(0.25)
Moderating Adverbs	1.76	1.36
	(1.72)	(1.28)

Note: USD-SD – Unique Sensory Detail and Spatial Detail

Figure 1. *Residualized Means of Interaction of Veracity Conditions and Indicators*



In order to examine how individual SA variables differed as a function of veracity condition and participant's language, we computed an overall Language (3) by Veracity Condition (2) MANOVA using the 6 SA variables as dependents. The main effect of Veracity Condition was significant, $\lambda = .874$, $F(6, 184) = 4.422$, $p < .000$, $\eta_p^2 = .126$. The main effect of Language was also significant, $\lambda = .724$, $F(12, 368) = 5.370$, $p < .000$, $\eta_p^2 = .149$. There was no interaction of Language and Veracity Condition, $\lambda = .899$, $F(12, 368) = 1.670$, $p = .074$, $\eta_p^2 = .051$.

To follow up the significant Veracity Condition main effect, we collapsed across languages and computed logistic regressions using Veracity Condition as the dependent variable and SA variables as covariates on the filtered data, using backward conditional exclusion criteria, in order to clarify which SA variables differentiated truthful and false statements. The final equation included three SA variables, USD-SD, Extraneous Information, Equivocation and accounted for 68.7 % overall correct classification of cases (Table 2). USD-SD, Extraneous Information, and Equivocation significantly differentiated true statements from deceptive statements.

Table 2. *Final results of logistic regressions, separately for written statements and investigative interviews*

Final Model	Overall Correct	False	False	Variables In	B	SE
Chi-Square	Classification	Positive	Negative			
	(%)	(%)	(%)			
$\chi^2(3, 163) = 34.172$, $p < .001$	68.70	14.72	16.56	USD-SD	1.035	.258

Extraneous Information	-.486	.243
Equivocation	-.366	.195

Note: USD-SD – Unique Sensory Detail and Spatial Detail

Posthoc analyses: Gender differences

According to Suckle-Nelson et al. (2010), women who responded deceptively were more aware of the need to keep their statement short and careful than were men who responded deceptively. Although Suckle-Nelson et al. did not use SA, it was possible that, regardless of language, gender matters. Thus, gender was tested in the current study by conducting an overall MANOVA using Language (3), Veracity Condition (2), and Gender (2 as factors on the SA variables. There was no significant effect of Gender, indicating that gender did not moderate the effects reported earlier.

Discussion

We examined whether the SA features that had been tested in the same language groups in the previous study (Matsumoto & Hwang, in press) would still reliably differentiate truthful statements from false statements in a different experimental and crime context from the previous study. The findings of the study supported the hypothesis that the SA features would differentiate truths from lies across languages and gender. Specifically, the categories USD-SD, Extraneous information, and Equivocation were significant indicators of veracity in multiple language groups. The results indicated that participants tended to write details, such as recalling particular scents, background noises or sounds, or locations, etc. and to provide information directly relevant to the incident than liars when delivering truths in comparison to deceptive statements. This finding is consistent with the recent literature that tested the function of SA in eyewitnesses' statements about a crime (Matsumoto & Hwang, in press).

Some cautions, however, need to be exercised in interpreting the results of the study. First, the study tested one type of crime, which was a mock crime of theft. The results may vary with other types of crimes. Matsumoto and Hwang (in press) reported that people tended to perceive crimes such as a hit-run crime similarly across language groups, and their findings vis-à-vis the function of SA were similar to the ones from the current study. Thus, it is possible that the SA categories examined in both studies in the same three language groups might be reliably applicable at least for the cases of mock crime and hit and run crime. Yet, there are still many possibilities and flexibility for the SA method to play a different role for other types of crimes, and these should be studied in the future. Second, we tested participants who had no previous experience with actual investigative situations; however, the reported findings might be different with people who were already exposed to similar or real investigative contexts. Thus, the finding is limited to people who are relatively naïve about investigative contexts (although there is an alternative possibility that people have indirect experiences through media or movies). Third, the gender ratio of one of the three groups was not perfectly equivalent as the Chinese group had relatively less males.

Despite of these limitations, this study makes several important contributions. First, it extended the previous study by Matsumoto and Hwang (in press) to a different context. Also, the current study replicated previous findings from many valuable studies that have examined the various SA strategies with different categories (Porter & Yuille, 1996; Vrij, 2000). Our finding added to the scientific evidence concerning the reliability of SA in distinguishing truthful statements from lying ones across different crime contexts.

Second, the reported findings are meaningful because the data were derived from the context in which the participants had to actually commit the criminal act although it was an experimental situation. Although we need further evidence ideally based on data from actual criminals or witnesses in order to test the real usability of SA in detecting deception, participants highly self-rated that the stakes of the conditions (punishment and reward), which means that the participants took stealing the \$200 check seriously and got nervous about the act. The significant SA features tested in the study can be possibly applicable to the real investigative contexts. Our findings increased the possibility of SA as a pragmatic method in distinguishing truths from lies in statements and guided which particular SA categories should be more paid attention in using SA as a constructive method in deception detection when analyzing written statements provided by suspects or witnesses. As one of the customary or conventional processes of investigation, SA could be a valuable aid in making the investigation process effective.

Third, the study indicated that SA is applicable across at least three languages (English, Chinese, Spanish) regardless of gender. This result is crucial because at least in the U.S., crimes became globalized and the number of immigrants has increased overtime. Dealing with non-English speakers and their statements in investigative contexts is not surprising or rare anymore. Law enforcement officers or interrogators who may have to deal with written statements or use them as a source of interviews can possibly utilize the SA method with non-English speakers once the officers obtain the analytic skills. Considering that the major immigrant groups speak Spanish or Chinese, not only in the U.S., but around the world, the SA approach would be pragmatic to use for particularly those three language groups regardless of gender as well as for bilingual speakers among the three languages.

Future studies will need to examine other languages (Arabic, French, etc.) using the SA method. Also, it would be interesting to test whether the current findings would vary depending on other different types of crime and levels of stakes. The current study examined whether basic factors such as language and gender could affect the efficacy of SA in deception detection. However, there must be other elements that should be examined in order to distinguish truthful statements from the lying ones. More varieties of research on SA would be desirable. Ideally collecting the SA data from people, not only born and raised, but also currently living in that country and using the first language would verify the pure usability of analyzing linguistic information across different languages.

References

- Adams, S. H. (1996). Statement analysis: What do suspects' words really reveal? *FBI Law Enforcement Bulletin*, 65, 12-20.
- Adams, S. H. & Jarvis, J. P. (2006). Indicators of veracity and deception: An analysis of written statements made to police. *Speech, Language, and the Law*, 13, 1-22.
- Christie, R. (1970). Scale construction. In R. Christie & F. L. Geis (Eds.), *Studies in Machiavellianism* (pp. 10-34). New York: Academic Press.
- Colwell, J., Hiscock-Anisman, C., & Fede, J. (2013). Assessment criteria indicative of deception: An example of the new paradigm of differential recall enhancement. In B.S.
- Colwell, K., Hiscock-Anisman, C. K., Memon, A., Taylor, L., & Prewett, J. (2008). Assessment Criteria Indicative of Deception (ACID): an integrated system of investigative interviewing and detecting deception. *Journal of Investigative Psychology and Offender Profiling*, 4, 167-180.
- Connelly, S., Allen, M. T., Ruark, G. A., Kligyte, V., Waples, E. P., Leritz, L. E., & Mumford, M. D. (2006). Exploring content coding procedures for assessing truth and deception in verbal statements, Year 3, Cumulative Final Report. Norman, OK: University of Oklahoma.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129, 74-118.
- Duran, N. D., Hall, C., McCarthy, P. M., & McNamara, D. S. (2010). The linguistic correlates of conversational deception: Comparing natural language processing technologies. *Applied Psycholinguistics*, 31, 439-462.
- Frank, M. G. & Svetieva, E. (2013). Deception. In D. Matsumoto, M. G. Frank & H. S. Hwang (Eds.), *Nonverbal Communication: Science and Applications* (pp. 121-144). Newbury Park, CA: Sage.
- Harlow, D. (2014). Statement analysis: the truth lies in the words. Property casualty 360. <http://www.propertycasualty360.com/2014/08/29/statement-analysis-the-truth-lies-in-the-words>
- Hauch, V., Blandon-Gitlin, I., Masip, J., & Sporer, S. L. (2012, 23-27 April 2012). *Linguistic cues to deception assessed by computer programs: A meta-analysis*. Paper presented at the Proceedings of the EACL 2012 Workshop on Computational Approaches to Deception Detection, Avignon, France.
- Johnson, M. K. (1988). Reality monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General*, 117, 390-394.
- Johnson, M. K. & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.
- Leo, R. A. (2008). *Police interrogation and American justice*. Harvard University Press. ISBN 978-0-674-02648-3.

- Matsumoto, D., Hwang, H. C., & Sandoval, V. A. (2013). Ethnic similarities and differences in linguistic indicators of veracity and lying in a moderately high stakes scenario. *Journal of Police and Criminal Psychology*. doi: 10.1007/s11896-013-9137-7
- Matsumoto, D. & Hwang, H. D. (in press). Emotional reactions to crime across cultures. *International Journal of Psychology*.
- Masip, J., Bethencourt, M., Lucas, G., Sanchez-San Segundo, M., & Herrero, C. (2012). Deception detection from written accounts. *Scandinavian Journal of Psychology*, 53, 103-111.
- Masip, J., Sporer, S.L., Garrido, E. & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime & Law*, 11, 99-122.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29, 665-675.
- Porter, S. & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior*, 20, 443-459.
- Tsai, J. L., Ying, Y.-W., & Lee, P. A. (2000). The meaning of "being Chinese" and "being American": Variation among Chinese-American young adults. *Journal of Cross-Cultural Psychology*, 31, 302-332.
- Porter, S., Birt, A. R., Yuille, J. C., & Lehman, D. R. (2000). The negotiation of false memories: The influence of interviewer and rememberer characteristics on memory distortion. *Psychological Science*, 11, 513-516.
- Porter, S. & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior*, 20, 443-459.
- Rabon, D. (1994). *Investigative Discourse Analysis*. Durham, NC: Carolina Academic Press.
- Ruby, C. L. & Brigham, J. C. (1997). The usefulness of the criteria-based content analysis technique in distinguishing between truthful and fabricated allegations: A critical review. *Psychology, Public Policy, and Law*, 3, 705.
- Rudacille, W. C. (1994). *Identifying lies in disguise*. Dubuque, IA: Kendall-Hunt.
- Schafer, J. R. (2007). *Grammatical differences between truthful and deceptive written narratives*. Unpublished doctoral dissertation.
- Schelleman-Offermans, K., & Merckelbach, H. (2010). Fantasy proneness as a confounder of verbal lie detection tools. *Journal of Investigative Psychology and Offender Profiling*, 7, 247-260.
- Snyder, M. (1974). Self-Monitoring of Expressive Behavior. *Journal of Personality and Social Psychology*, 30, 526-537.

- Spence, K., Villar, G., & Arciuli, J. (2012). Markers of deception in Italian speech. *Frontiers in Psychology*, 3, 453. doi: 10.3389/fpsyg.2012.00453
- Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology*, 20, 421-446.
- Suckle-Nelson, J. A., Colwell, K., Hiscock-Anisman, C., Florence, S., Youschak, K. E., and Duarte, A. (2010). Assessment criteria indicative of deception (ACID): replication and gender differences. *Open Criminol. J.* 3, 23–30.
- Undeutsch, U. (1954). Die Entwicklung der gerichtropsychologischen Gutachtertätigkeit. In A. Wellek (Ed.), Bericht über den 19. Kongress der Deutschen Gesellschaft für Psychologie (pp. 1132-154). Göttingen: Verlag für Psychologie.
- Undeutsch, U. (1989). The development of statement reality analysis. In J. C. Yuille (Ed.), *Credibility Assessment* (pp. 101-119). New York: Kluwer Academic/Plenum Publishers.
- Vrij, A. (2000). *Detecting lies and deceit*. Chichester, England: Wiley.
- Vrij, A. (2007). Criteria based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, 11, 3-41.
- Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities*, 2nd ed., Chichester, England: John Wiley & Sons.
- Weintraub, W. (1989). *Verbal behavior in everyday life*. New York: Springer.
- Willén, R.M., & Strömwall, L.A. (2012). Offenders' uncoerced false confessions: A new application of statement analysis? *Legal and Criminological Psychology*, 17, 346-359. DOI: 10.1111/j.2044-8333.2011.02018.x