

# Miami University deception detection database

**E. Paige Lloyd, Jason C. Deska, Kurt Hugenberg, Allen R. McConnell, Brandon T. Humphrey & Jonathan W. Kunstman**

**Behavior Research Methods**

e-ISSN 1554-3528

Behav Res

DOI 10.3758/s13428-018-1061-4



**Your article is protected by copyright and all rights are held exclusively by Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**



## Miami University deception detection database

E. Paige Lloyd<sup>1</sup> · Jason C. Deska<sup>2</sup> · Kurt Hugenberg<sup>3</sup> · Allen R. McConnell<sup>4</sup> ·  
Brandon T. Humphrey<sup>4</sup> · Jonathan W. Kunstman<sup>4</sup>

© Psychonomic Society, Inc. 2018

### Abstract

In the present work, we introduce the *Miami University Deception Detection Database* (MU3D), a free resource containing 320 videos of target individuals telling truths and lies. Eighty (20 Black female, 20 Black male, 20 White female, and 20 White male) different targets were recorded speaking honestly and dishonestly about their social relationships. Each target generated four different videos (i.e., positive truth, negative truth, positive lie, negative lie), yielding 320 videos fully crossing target race, target gender, statement valence, and statement veracity. These videos were transcribed by trained research assistants and evaluated by naïve raters. Descriptive analyses of the video characteristics (e.g., length) and subjective ratings (e.g., target attractiveness) are provided. The stimuli and an information codebook can be accessed free of charge for academic research purposes from <http://hdl.handle.net/2374.MIA/6067>. The MU3D offers scholars the ability to conduct research using standardized stimuli that can aid in building more comprehensive theories of interpersonal sensitivity, enhance replication among labs, facilitate the use of signal detection analyses, and promote consideration of race, gender, and their interactive effects in deception detection research.

**Keywords** Stimulus set · Video database · Lie detection

Detecting deception is essential for interpersonal, intergroup, and societal functioning. Indeed, lying is a surprisingly common event, with the average individual reporting lying once or twice each day (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996). Misreading deception as benign can lead to lost resources, feelings of foolishness, distrust, and even dissolution of close relationships (e.g., Belot, Bhaskar, & van de Ven, 2010; Carton, Kessler, & Pape, 1999; Planalp & Honeycutt, 1985). Conversely, wrongful assertions of deceit can be embarrassing, heighten interpersonal tension, undermine intimacy, and terminate existing relationships. Despite the importance of accurate deception detection, people are only slightly better than chance at discriminating between lies and truths in deception detection tasks (i.e., 54% accuracy, with 50% being

chance; Bond & DePaulo, 2006, 2008; Vrij, Edward, Roberts, & Bull, 2000).

Many moderators and predictors of lie detection accuracy have been investigated, including expertise (e.g., police vs. laypeople; Bogaard, Meijer, Vrij, & Merckelbach, 2016; Mann, Vrij, & Bull, 2004), cognitive resources (e.g., spontaneous vs. deliberative judgments of deception; Albrechtsen, Meissner, & Susa, 2009; ten Brinke, Stimson, & Carney, 2014), and the use of specific cues (for a meta-analysis, see DePaulo et al., 2003). Furthermore, beyond detecting lies, research has documented response biases in these decisions. For example, people often demonstrate a “truth bias,” wherein they generally rely on the “truth” response more than the “lie” response (Levine, Park, & McCornack, 1999).

The considerable interest in understanding lie detection has led to an assortment of research paradigms and stimuli. For example, scholars have used relatively low-stakes lies about friendships and social events (e.g., Bond, Omar, Mahmoud, & Bonser, 1990; DePaulo & Rosenthal, 1979; Evans, Michael, Meissner, & Brandon, 2013) as well as more high-stakes lies about committing mock crimes (e.g., Frank & Ekman, 1997; ten Brinke et al., 2014) and falsehoods delivered by convicted criminals (e.g., Kassin, Meissner, & Norwick, 2005; ten Brinke & Porter, 2012). However, to the authors’ knowledge, there are few open-access databases of videos for use in deception

✉ E. Paige Lloyd  
emilypaigelloyd@gmail.com

<sup>1</sup> University of Denver, 2155 S. Race St., Denver, CO 80433, USA

<sup>2</sup> University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3, Canada

<sup>3</sup> Indiana University, 1101 E. 10th St., Bloomington, IN 47405, USA

<sup>4</sup> Miami University, 90 North Patterson Avenue, Oxford, OH 45056, USA

detection research, and the lack of standardized stimuli can lead researchers to fashion their own experimental materials.

It is evident that creating such materials is daunting and may impede research, but the lack of standardization in lie detection stimuli also contributes to difficulty in comparing findings across labs and inhibits the development of unified theories of lie detection. Furthermore, because stimulus creation is time-intensive, many deception detection findings are drawn from small samples of stimuli (often fewer than 20 stimuli) in which important target characteristics (e.g., race, gender) are neither controlled nor considered. This lack of stimulus control becomes more concerning when we consider that target characteristics play a larger role in deception detection accuracy and response biases than do perceiver characteristics (Bond & DePaulo, 2008; Levine et al., 2011). Thus, the deception detection literature would benefit from a large, open-access database of standardized deception detection videos featuring a number of important target characteristics that are of interest to researchers.

Accordingly, we introduce the *Miami University Deception Detection Database* (MU3D), a free resource containing 320 videos of Black and White females and males telling positive and negative truths and lies. We first briefly review the commonly used stimuli in the lie detection literature, noting the potential strengths and weaknesses of each approach. Next, we discuss how the MU3D was created, developed, and transcribed. Finally, we present data on the objective characteristics of the videos, as well as by-stimulus analyses of raters' evaluations of the videos and of the targets featured in the videos. These analyses indicate consistency in the video characteristics (e.g., length), adequate interrater reliability, and meaningful variability in the subjective ratings. Additionally, the results from correlation and multiple regression analyses illuminate novel insights and suggest fertile avenues for future research. Overall, the MU3D can help advance research in deception detection and additionally can offer stimuli that will be of interest to social scientists who study phenomena beyond lie detection (e.g., intergroup relations).

## Brief review of existing stimuli

In our review of the literature, we discuss two of the most common categories of deception detection stimuli. Then we discuss the similarities and differences between these stimulus categories and the MU3D, with special attention given to the relative strengths and limitations of our new database.

**Opinions, personal facts, and person descriptions** Among the most common deception detection paradigms are those in which stimuli are generated from individuals who tell truths or lies about personal information (e.g., life events, relationships, attitudes). In DePaulo et al. (2003) meta-analysis on cues to deception, slightly less than half (42.50%) of the 120 studies

analyzed used as their stimuli truth and lie videos concerned with the targets' attitudes, personal relationships, or facts. In attitude or opinion paradigms, participants complete measures assessing their opinions on social issues (e.g., "Should convicted cold-blooded murderers be executed?") and later are recorded speaking truthfully or deceitfully about a particular opinion (often about the issue that they felt most strongly about; see, e.g., Frank & Ekman, 1997; Leal, Vrij, Mann, & Fisher, 2010). In personal-fact paradigms, participants accurately or falsely report on personal facts or events. For example, in the true or false alibi paradigm, participants are asked to report on their actual or fabricated whereabouts at a specified date and time (e.g., "where were you last Saturday night from 7:00 p.m. to 10:00 p.m."; Evans et al., 2013). In other research using personal-fact paradigms, participants report on stressful life events (e.g., car accident, death of loved one) that are either fictitious or true (e.g., Porter, Yuille, & Lehman, 1999; Sporer, 1997). Finally, in personal-description paradigms (e.g., DePaulo & Rosenthal, 1979), participants tell truths and lies about their social relationships (e.g., "describe a person you dislike as if you like them").

**Cheating and mock crimes** Cheating and mock-crime paradigms are also frequently used in the deception literature, comprising 13.30% of the studies reported in DePaulo and Bond's (2003) meta-analysis. Within this category, the methods used are quite diverse. Levine (2007) created a database of 111 cheater paradigm deception detection videos, which are accessible to scholars upon request. In this database, participants played a trivia game in which cheating would benefit their performance and their likelihood of rewards. All participants were provided with the opportunity to cheat, and they freely chose whether to cheat or to follow the rules. Later, participants were questioned about their behavior. Liars were operationalized as those who cheated and later lied about cheating, whereas truth tellers were those who followed the rules and reported doing so (all other participants demonstrated a combination of truthful and deceitful behavior). Although there are 111 total videos in the database, only 22 individuals cheated and lied about cheating. Thus, much of the subsequent work using this database has used a final set of 44 videos (all 22 cheaters and a matched selection of 22 truth tellers) as stimuli (e.g., Levine, Shaw, & Shulman, 2010). Levine and colleagues have since developed several other cheater paradigm video sets, available upon request. These video sets feature similar cheater paradigms with alterations to the specific procedure (e.g., manipulating whether a partner instigates cheating or altering postevent interview questions; see Levine, Blair, & Clare, 2014; Levine et al., 2014; Levine, Shulman, Carpenter, DeAndrea, & Blair, 2013).

Other cheating paradigms use still image stimuli instead of videos. For example, Verplaetse, Vanneste, and Braeckman (2007) captured images of 112 individuals engaging in a

prisoner's dilemma game with a partner. Still images were taken just before individuals either defected (i.e., cheated their partner for selfish reward) or cooperated (i.e., made a choice that equally benefited themselves and their partner). Of these 112 target images, 26 (13 cooperative and 13 uncooperative), selected on the basis of image quality and target characteristics (e.g., matched numbers of men and women), are commonly used in subsequent work (e.g., Shoda & McConnell, 2013).

In mock-crime paradigms adapted from work in criminal justice (Kircher, Horowitz, & Raskin, 1988), individuals are given the opportunity to commit a crime (e.g., steal money) and later are interrogated about the episode (e.g., Frank & Ekman, 1997; ten Brinke et al., 2014). Truth tellers are those individuals who did not commit the crime and accurately reported their innocence to the interrogator, whereas liars are those individuals who committed the crime but claimed innocence. For example, ten Brinke et al. (2014) developed 12 (six genuine videos and six deceptive videos, with equal numbers of men and women in both groups) high-quality mock-crime stimulus videos, which are available to researchers upon request. In developing their stimuli, individuals were randomly assigned to steal or not to steal \$100 from an envelope. They were told that if they could convince the experimenter they had not stolen the money (regardless of their innocence), they would earn \$100 and be entered in a lottery for an additional \$500.

**Stimulus strengths and weaknesses** There are strengths and weakness of both types of lie detection stimuli. Most notably, the *opinions*, *personal-fact*, and *person-description* paradigms typically offer greater experimental control. For example, because all people have some relationships or facts about which they can tell truths or lies, experimenters can ensure that any given individual tells a truth or tells a lie rather than relying on the individual's self-selected behavior (e.g., those who happen to steal money and lie about it). In addition, because personal facts, opinions, and relationships are available to all, it is relatively easier to generate such stimuli. However, this experimental control may come at the expense of generalizability to high-stakes or criminal deception. Indeed, no actual crimes or cheating takes place in these scenarios, and the deception motivation is external (i.e., an experimenter instructing the participant to lie) rather than internal (i.e., a suspect attempting to avoid conviction). Thus, these lies differ in both magnitude and locus (i.e., externally vs. internally motivated). Because of these differences, some researchers have argued that "instructed lies" (e.g., stimuli generated from opinions, personal facts, person descriptions, or even mock-crime scenarios in which cheating behavior is randomly assigned rather than perpetrated by choice) may not capture situations involving genuine, internally motivated lies (Levine, 2018). Although we agree that there likely are notable differences between the experiences of committing a crime or infraction and lying about it

freely and of cheating or lying because instructed to do so, we argue that both types of lies are forms of deception worthy of investigation. For example, if a boss instructs a sales employee to lie to a customer in service of making a sale, this may be externally motivated but still ecologically meaningful.

Also, trade-offs are involved for the cheating and mock-crime paradigms. Although these paradigms can offer mundane realism because of their relatively high-stakes deception situations, the lack of experimental control may introduce potential third variables (e.g., individual differences that lead one person to lie may also influence how people tell lies or their general efficacy at lying). Indeed, the experimenter in cheating and mock-crime situations does not typically have control over who decides to defect against others in social dilemma situations (e.g., Verplaetse et al., 2007) or who chooses to commit crimes and then also chooses to tell truths or lies about their actions (e.g., Levine, 2007; cf. ten Brinke et al., 2014). Thus, although these paradigms feature lies that are relatively high-stakes and internally motivated, the choice to commit an initial infraction and the choice to lie are frequently confounded. Furthermore, such paradigms rarely offer a within-stimulus manipulation of truths and lies—because a person either stole or did not steal the money, the same individual cannot tell both a truth and a lie about not stealing the money.

**Advantages of the present approach** In developing the MU3D, we used the personal-description paradigm in order to maximize experimental control and generalizability to low-stakes, everyday forms of deception. Indeed, most lies are primarily low-stakes and involve common interpersonal interactions rather than high-stakes situations, such as criminal behavior (DePaulo et al., 1996). Even these relatively low-stakes deceptions impact relationship success and well-being (Carton et al., 1999; DePaulo et al., 1996; Planalp & Honeycutt, 1985). We acknowledge that this choice may limit some of the potential uses of MU3D, but at the same time, the experimental control that personal-description paradigms offer, coupled with the pervasiveness of low-stakes lies, makes this a reasonable starting point.

Furthermore, by focusing on a context in which statement control is possible, the personal-description paradigm that we adopted allowed us to manipulate statement content and valence, as well as to manipulate important (yet understudied) target-level individual differences. Specifically, in the MU3D, each target participant provided four statements, two lies and two truths, crossed orthogonally with valence, about their personal relationships. In other words, each target participant told a positive truth, a positive lie, a negative truth, and a negative lie, so that statement content and valence are crossed factorially.

This stimulus control provides many benefits for researchers. For example, when telling a truth about someone a target likes and when lying about liking someone a target dislikes, the content of the statements can be quite similar, allowing researchers to examine lie detection involving

stimuli with similar verbal content. Furthermore, whereas mock-crime and cheating paradigm statements are always about negative scenarios (e.g., stealing), the personal-description paradigm allows for valence to be fully crossed with veracity, providing the possibility to study how lie detection differs across valences. For example, are positive lies more difficult to detect than negative lies? Is there a stronger truth bias (i.e., especially heavy reliance on “truth” responses) for positive than for negative statements? Questions such as these are important, but they are unanswerable using most existing stimulus sets. Finally, the personal-description paradigm allows researchers to investigate key individual differences in both the targets and perceivers. In many mock-crime and cheating paradigms, individual targets choose whether to act deceitfully. Thus, the targets either tell the truth or a lie, but rarely both. Thus, stable individual differences that correspond with deceptive action are confounded with veracity in these stimuli. By having targets tell both truths and lies, the targets act as their own control, creating within-subjects manipulations of truths and lies.

Third, a particular strength of the MU3D is that it includes 320 stimulus videos, which we believe constitutes the largest set of stimuli available to lie detection researchers. This relatively large quantity of stimuli provides at least three advantages. First, the greater number of videos increases external validity and allows researchers to incorporate target-level variables, such as sex and race, into research designs. As we noted above, much existing deception detection research overlooks important features of the target, despite the large role that target characteristics play in deception detection accuracy and biases (Bond & DePaulo, 2008). By considering targets' age, race, ethnicity, gender, and even college major, researchers can begin to investigate target features that may impact deception detection, or at the very least can control for target-level variance. The inclusion of both Black and White targets in the database may also attract interest from researchers outside the deception detection literature. For example, intergroup researchers might find knowledge of own-race advantages or biases in deception detection informative (Lloyd, Hugenberg, McConnell, Kunstman, & Deska, 2017). Moreover, simply having videos of Black and White female and male targets talking on camera about social relationships could be of value to researchers and educators working in the realm of intergroup dynamics.

Second, the large number of stimuli enables researchers to conduct by-stimulus analyses (in addition to the by-participant analyses that are typical in many psychology studies), providing new possibilities for understanding lie detection. Indeed, the traditional approach in psychological research is to have a large number of participants view a small number of stimuli. Through by-stimulus analyses, researchers can instead begin to ask questions about characteristics of the lie teller or truth teller. When conducting by-stimulus analyses, researchers

may not even need to invest in additional participant or human subject recruitment. For example, researchers could use the Linguistic Inquiry and Word Count application (LIWC; Pennebaker, Booth, Boyd, & Francis, 2015) to compare whether the verbal contents of lies and truths differ across the 320 videos included in the dataset. For example, truths or lies may feature different types of words, frequencies of words, or linguistic complexity.

Finally, having a large number of database videos encourages researchers to conduct *signal detection analyses* (Green & Swets, 1966; Macmillan & Creelman, 2005), providing greater precision about the mechanisms underlying lie detection. Historically, researchers have typically focused on accuracy as the key metric of lie detection. Signal detection approaches allow researchers to avoid confounding the *ability to discriminate* between truths and lies (i.e., sensitivity) with the *tendency to favor* one response over another (i.e., response bias). Few studies in the lie detection literature have used signal detection analyses (cf. Albrechtsen et al., 2009; Lloyd et al., 2017), a shortcoming that is likely due, at least in part, to studies using relatively few target stimuli, which makes conducting signal detection analyses problematic. Thus, having a database with a larger number of stimuli will provide researchers with more sophisticated inferential tools, and accordingly, with better insights into what underlies accuracy in lie detection studies.

## Database creation

### Overview of database creation and ratings

All videos in the MU3D are available for download to academic researchers at no cost from <http://hdl.handle.net/2374.MIA/6067>. A codebook provides researchers with additional information about each video (e.g., trustworthiness ratings, anxiety ratings, length of video, transcriptions of videos), as well as information about the targets featured in the videos (e.g., attractiveness ratings, self-reported age, and self-reported race). Creation of the MU3D involved two waves of data collection: a *stimulus generation* wave and a *stimulus rating* wave. Each wave involved different groups of participants. The stimulus generation was closely modeled on previous personal-description paradigms (e.g., Bond et al., 1990; DePaulo & Rosenthal, 1979). Participants were instructed to talk about (1) a person they liked (positive truth), (2) a person they disliked (negative truth), (3) the person they liked as if they disliked that person (negative lie), and (4) the person they disliked as if they liked that person (positive lie). Thus, statement valence (positive vs. negative) and veracity (truth vs. lie) were crossed within targets. After the stimuli were recorded, they were transcribed. The final database included 320 videos in a fully factorial mixed design: 2 (Race: Black vs. White;

between subjects)  $\times$  2 (Sex: male vs. female; between subjects)  $\times$  2 (Valence: positive vs. negative; within subjects)  $\times$  2 (Veracity: honest vs. dishonest; within subjects).

During the stimulus rating wave, a separate sample of participants (hereafter referred to as *raters*) viewed the videos and engaged in two tasks. First, raters attempted to discriminate truths from lies for each video viewed. Second, they evaluated each of the targets in the videos on attractiveness, trustworthiness, and anxiety.

### Wave 1: Stimulus generation

**Target participants** A total of 112 (33 Black female, 27 Black male, 25 White female, and 27 White male) students and staff members were recruited from Miami University's campus, who ranged in age from 18 to 26 years ( $M_{age} = 20.13$ ,  $SD_{age} = 1.56$ ). Participants volunteered without reimbursement, received partial course credit, or were paid \$20.

**Apparatus** The videos were recorded using a c525 Logitech HD Webcam with a video resolution of  $1,280 \times 720$  and a frame rate of 30 fps, connected to a computer to record all videos and audio using its built-in microphone.

**Stimulus collection** Upon entering the laboratory, the target participants were seated in a private, webcam-equipped cubicle and consented to participate. They then engaged in the video-recorded lie detection task. The experimenter began each recording with a prompt, left the cubicle, and returned 45 s later to stop the recording. For the first video, the experimenter said: "Please describe a person you know who you truly like, talk about why you like that person, and describe their positive qualities." All participants were asked not to use the name of the person described, but instead to use only the pronoun "him," "her," or "they." The experimenter started a video recording and a 45-s timer before exiting the cubicle. After the 45 s elapsed, the experimenter administered the second prompt: "Now, describe the same person you just spoke about, but this time *lie* and describe that person as if you dislike them, talk about why you dislike that person, and describe their negative qualities." After 45 s, this procedure was repeated again with the following prompts: "Please describe a person you know who you truly dislike, talk about why you dislike that person, and describe their negative qualities" and "Now, describe the same person you just spoke about, but this time *lie* and describe that person as if you like them, talk about why you like that person, and describe their positive qualities."

This procedure resulted in the generation of four relatively low-stakes lie detection videos per participant, fully crossing valence (negative vs. positive) with

veracity (truths vs. lies). All participants responded to the prompts in the same order,<sup>1</sup> were urged to be convincing, and were asked to speak for the entire 45 s. After responding to all four prompts, participants completed a demographics questionnaire assessing age, race, ethnicity, gender, and major or affiliation with the University, before being thanked, compensated (if applicable), and debriefed. After debriefing, participants were asked to separately provide consent for the researchers to keep and edit their videos, and for the researchers to use their videos in future research. All participants signed both video consent forms enabling their videos to be edited and shared as a research resource.

**Stimulus selection** In total, 448 videos were created (four videos each for 112 target participants). We selected 20 targets from each of the four target categories (i.e., Black female, Black male, White female, White male), yielding a total of 320 selected videos. These targets were selected primarily on the basis of a priori inclusion criteria: Participants responded to all four prompts, spoke for at least 20 s in each of the four videos, followed directions (e.g., spoke about a friend as opposed to an actor or a political figure), did not disrupt the camera with excessive movement (e.g., banging on the table), and remained in frame throughout all four videos. When more than 20 targets met the a priori inclusion criteria, we selected the 20 targets with the best video quality, and when the video quality was not visibly different, targets were chosen on the basis of random selection.<sup>2</sup>

**Stimulus preparation** The 320 selected videos were cropped to remove the first few seconds (i.e., the period when the experimenter left the cubicle) and the tone that signaled the end of 45 s. The videos were cropped so that they concluded at the end of a complete sentence or thought. Videos in which participants did not speak for the full 45 s were cropped to eliminate periods of silence near the end of the video. The exact length (and other information) for each video is included in the MU3D codebook.

<sup>1</sup> Because the prompts were always given in the same order, order effects could be a concern. For example, participants could have been more convincing when telling positive lies (Video 4) than negative lies (Video 2) or than telling negative truths (Video 3) than positive truths (Video 1) because they had previously practiced telling a lie or the truth. This pattern of results would best be represented by a valence by veracity interaction (i.e., greater "truth" responding to positive than to negative lies, but greater "truth" responding to negative than to positive truths). Although order effects cannot be completely ruled out, given the present study design, we conducted a 2 (Valence)  $\times$  2 (Veracity) ANOVA on "truth" responding. We found no evidence of such a valence by veracity interaction [ $F(1, 316) = 2.10$ ,  $p = .148$ ].

<sup>2</sup> Only the 20 selected targets from each category were transcribed, rated in Wave 2, and included in the MU3D codebook. However, researchers might be interested in using the unselected targets in their own research. We are happy to share these videos as well, upon request.

## Inter-wave: Database transcriptions

**Transcribers** Six trained undergraduate research assistants (five female, one male) transcribed the verbal content from the 320 selected deception detection videos.

**Transcription procedure** During an eight-week period, each research assistant was assigned six to eight videos each week to transcribe. Transcribers had access to both video and audio content while generating their transcriptions. The six transcribers and project PI met weekly to discuss any transcription issues or questions (e.g., statements that were difficult to understand, transcribers questioning whether to report a noise generated by the target as a verbalization). If issues were reported, the entire group watched the video until unanimous agreement resolved the concern. Transcriptions include all word and word-like vocalizations (e.g., um, uh) but not other vocalizations (e.g., coughs, throat clearing, laughs).

## Wave 2: Deception detection and stimulus ratings

**Participants** A total of 405 participants, recruited from Amazon's Mechanical Turk, attempted to accurately detect deception from the videos and provided subjective ratings of the videos in exchange for monetary compensation (\$1.50). Our goal was to collect as many high-quality ratings as possible, to maximize reliable mean values while not exceeding our available funding resources. The raters included 203 men and 199 women (three did not disclose their gender) and were primarily White (338 White, 28 Black, 26 Asian, seven bi- or multiracial, three American Indian or Alaska Native, one Native Hawaiian or Pacific Islander, one other, and one non-response) and ranged in age from 18 to 78 ( $M_{age} = 34.45$ ,  $SD_{age} = 11.34$ ).

**Procedure** The raters were randomly assigned to one of 20 video sets created from the 320 deception detection videos. Each video set included a negative truth, negative lie, positive truth, and positive lie from four different targets from each speaker demographic category (i.e., Black female, Black male, White female, White male). Thus, each rater evaluated 16 videos from the database (i.e., four Black female videos, four Black male videos, four White female videos, and four White male videos), but never saw a given target more than once. Each rater viewed the same number of lie and truth videos from each demographic category of speaker. Furthermore, each rater viewed an equal number of positive and negative statements from each demographic category of speaker. The 16 videos to be evaluated were presented one at a time in a randomized order. Progression through the study was controlled such that participants could not continue to the ratings until the duration of the video had elapsed. Following each video, raters responded to four questions in

a fixed order: "Is this person telling a truth or a lie?," "How attractive is this person?," "How trustworthy is this person?," and "How anxious is this person?" The first question was assessed via a forced-choice dichotomous selection of "Truth" or "Lie." The remaining three questions were assessed with scales ranging from 1 (*Not at all*) to 7 (*Extremely*).

## Results

### Video information

Detailed information about each video (as well as target level video ratings) is available in the codebook. Here, we report descriptive statistics about the videos. First, the videos averaged 35.73 s in length (Min = 24.33, Max = 43.60,  $SD = 3.49$ ). The lengths of the female targets' ( $M = 35.84$ ,  $SD = 3.52$ ) and male targets' ( $M = 35.61$ ,  $SD = 3.47$ ) videos were comparable,  $t(318) = .56$ ,  $p = .574$ ,  $d = 0.07$ , as were the lengths of videos by Black ( $M = 35.59$ ,  $SD = 3.30$ ) and White ( $M = 35.86$ ,  $SD = 3.68$ ) targets,  $t(318) = -0.69$ ,  $p = .490$ ,  $d = -0.08$ . Furthermore, the negative ( $M = 35.80$ ,  $SD = 3.34$ ) and positive ( $M = 35.66$ ,  $SD = 3.64$ ) videos were of similar length,  $t(318) = .37$ ,  $p = .713$ ,  $d = 0.04$ , as were the videos of lies ( $M = 35.54$ ,  $SD = 3.54$ ) and truths ( $M = 35.92$ ,  $SD = 3.44$ ),  $t(318) = -0.96$ ,  $p = .337$ ,  $d = -0.11$ .

Similar analyses were conducted on the mean word counts. The transcriptions of the videos averaged 106.69 words in length (Min = 46.00, Max = 160.00,  $SD = 23.48$ ). The analyses indicated that the female videos ( $M = 102.44$ ,  $SD = 24.09$ ) had fewer words than the male videos ( $M = 110.94$ ,  $SD = 22.12$ ),  $t(318) = -3.29$ ,  $p = .001$ ,  $d = -0.37$ , and that the videos of Black targets ( $M = 102.46$ ,  $SD = 24.38$ ) had fewer words than those of White targets ( $M = 110.92$ ,  $SD = 21.81$ ),  $t(318) = -3.27$ ,  $p = .001$ ,  $d = -0.37$ . However, the negative ( $M = 108.36$ ,  $SD = 22.54$ ) and positive ( $M = 105.01$ ,  $SD = 24.34$ ) videos were comparable in word count,  $t(318) = 1.28$ ,  $p = .202$ ,  $d = 0.14$ , as were the transcriptions of videos with lies ( $M = 107.57$ ,  $SD = 24.35$ ) and truths ( $M = 105.81$ ,  $SD = 22.62$ ),  $t(318) = 0.67$ ,  $p = .503$ ,  $d = 0.07$ . In the present work, we do not explore transcription content, but we anticipate that future researchers will investigate questions involving speech content and deception detection.

### Lie detection and video ratings

Below, we present descriptive analyses of the deception detection task and subjective video ratings at both the video ( $N = 320$ ) and target ( $N = 80$ ) levels. These are by-stimulus analyses because the raters viewed only a small subset of the available videos. A by-participant analysis was impractical here, given the number of missing observations that such an analysis would have entailed. Our



primary goals were to assess rater reliability, performance (i.e., truth bias and accuracy), and subjective judgments.

**Video-level analyses** We first calculated the interrater reliability of video attractiveness ( $\alpha = .87$ ), trustworthiness ( $\alpha = .63$ ), and nervousness ( $\alpha = .73$ ) ratings. We then calculated average attractiveness, nervousness, and trustworthiness ratings for each video across all participants who had viewed the video. We also computed the proportions of correct responses (i.e., accuracy) and the proportions of “truth” responses (i.e., truth proportion) from the perceivers of each video. Signal detection analyses were not feasible for video-level analyses. The calculation of sensitivity and bias (described in greater detail in the target-level analyses section below) necessitate the calculation of hits (i.e., correct identification of lies) and false alarms (i.e., incorrect labeling of truths as lies) for the stimulus of interest. However, a given video could only contribute either hits (i.e., lie videos) or false alarms (i.e., truth videos), making signal detection analyses impossible.

Table 1 presents descriptive statistics and correlations among the variables averaging across all videos in the stimulus set. Zero-order correlations indicated that accuracy and truth proportion were unrelated. Furthermore, accuracy was not associated with attractiveness, trustworthiness, or anxiousness. Thus, at the video level, it does not appear that attractiveness, trustworthiness, or anxiousness served as effective cues to deception detection. However, the proportion of “truth” responses was associated with trustworthiness and anxiousness, but not with attractiveness. In other words, the videos in which targets appeared relatively more trustworthy or less anxious were ascribed relatively more “truth” responses. Finally, and perhaps unsurprisingly, trustworthiness and anxiousness were negatively correlated.

We also conducted multiple regression analyses to explore the independent effects of attractiveness, trustworthiness, and anxiousness on both accuracy and truth proportion (see Table 2). Mirroring the results above, we observed no evidence that video-level ratings of attractiveness, trustworthiness, or anxiousness predicted accuracy. However, video-level ratings of attractiveness, trustworthiness, and anxiousness each uniquely contributed to the

**Table 1** Video-level analyses, including means, standard deviations, and correlations among the variables

Variable	Mean	SD	1	2	3	4
1. Accuracy	.52	.21				
2. Truth proportion	.59	.18	.03			
3. Attractive	4.08	0.58	.09	.09		
4. Trustworthy	4.16	0.52	.02	.83***	.21***	
5. Anxious	3.04	0.65	.03	-.65***	-.13*	-.64***

\*  $p < .05$ . \*\*\*  $p < .001$ .

**Table 2** Video-level multiple regression analyses regressing accuracy and truth proportion on attractiveness, trustworthiness, and anxiousness

Predictors	Outcome variables
	Accuracy
	$\beta$
Attractive	.08
Trustworthy	.05
Anxious	.08
	Truth proportion
	$\beta$
Attractive	-.09**
Trustworthy	.72***
Anxious	-.20***

\*\*  $p < .01$ . \*\*\*  $p < .001$

truth judgments. Videos featuring relatively more anxious and more attractive targets were less trusted (i.e., were ascribed fewer “truth” responses), whereas videos featuring more trustworthy targets were more trusted. The accuracy, truth proportion, and subjective ratings for specific videos are available in the video-level tab of the codebook.

**Target-level analyses** We first calculated the interrater reliability of target attractiveness ( $\alpha = .95$ ), trustworthiness ( $\alpha = .84$ ), and nervousness ratings ( $\alpha = .91$ ), and then we calculated the mean attractiveness, nervousness, and trustworthiness ratings for each target. Because all targets supplied both truthful and deceitful statements (allowing for calculation of both hits and false alarms; see below for more details), we could use signal detection analyses at the target level. Thus, in addition to accuracy and proportion of “truth” responses, we also calculated sensitivity ( $d'$ ) and criterion ( $c$ ). Sensitivity indexes the raters’ ability to discriminate targets’ truths from lies, whereas the criterion indexes the degree to which individual raters favored the “truth” response over the “lie” response. To calculate sensitivity and criterion, we first calculated the proportions of hits (i.e., correct identification of a lie) and false alarms (i.e., calling a truthful statement a lie) and, as is common in signal detection analyses, cells with proportions of 1 or 0 were replaced with .99 or .01, respectively (Macmillan & Kaplan, 1985). These proportions were then standardized, and sensitivity was calculated by subtracting the standardized false alarms from the standardized hits. Greater sensitivity ( $d'$ ) values indicate that raters were better able to distinguish truths from lies (i.e., better deception detection). Criterion was calculated by adding the standardized measures of hits and false alarms before dividing by  $-2$ . Thus, greater criterion values indicate more “truth” responses and fewer “lie” responses (i.e., a greater truth bias).

Table 3 presents descriptive statistics and correlations among the variables, averaging across all targets in the

**Table 3** Target-level analyses, including means, standard deviations, and correlations among the variables

Variable	Mean	SD	1	2	3	4	5	6
1. Sensitivity	.10	.38						
2. Criterion	.26	.37	-.01					
3. Accuracy	.52	.07	.99***	-.04				
4. Truth proportion	.60	.13	-.01	.99***	-.04			
5. Attractive	4.08	0.51	.29**	.03	.29**	.03		
6. Trustworthy	4.16	0.29	.06	.91***	.03	.91***	.11	
7. Anxious	3.04	0.51	-.04	-.75***	-.02	-.76***	-.08	-.73***

\*\*  $p < .01$ . \*\*\*  $p < .001$

stimulus set. As we would expect in any design with equal representation of truths and lies (i.e., 50% truths and 50% lies), sensitivity and accuracy ( $r = .99$ ) as well as criterion and truth proportion ( $r = .99$ ) were strongly correlated. However, as researchers employ methods that deviate from 50% truths and 50% lies, these correlations would be expected to shift. Thus, despite these strong correlations, we discourage readers from concluding that sensitivity and accuracy or criterion and truth proportion are interchangeable terms across situations or research paradigms. Zero-order correlations indicated that both accuracy and sensitivity were positively related to attractiveness, with more attractive targets being easier for raters to “read” correctly. Congruent with the video-level analyses, criterion and truth proportion were related to trustworthiness and anxiousness. Specifically, both truth proportion and criterion increased as targets were rated as less anxious or more trustworthy.

We then conducted multiple regression analyses to explore the independent effects of attractiveness, trustworthiness, and anxiousness on sensitivity, criterion, accuracy, and truth proportion (see Table 4). The results indicated that for sensitivity and accuracy, attractiveness was the only unique predictor. That is, relatively more attractive targets predicted greater deception detection among raters. Ratings of target trustworthiness and anxiousness again predicted truth proportion as well as criterion. Specifically, relatively more anxious targets elicited less “truth” responding from raters, whereas relatively more trustworthy targets elicited greater “truth” responding from raters. For the target-level mean ratings for accuracy, truth proportion, sensitivity, and criterion values, see the target-level database codebook.

### General discussion

The ability to detect deception is a valuable skill that predicts important life outcomes (e.g., Belot et al., 2010; Carton et al., 1999; Planalp & Honeycutt, 1985). For social scientists conducting research on this topic, having a large pool of freely available stimuli would benefit scholars who study deception,

intergroup relations, and social perception more broadly. At present, some valuable resources have been made available by ten Brinke et al. (2014) and by Levine (2007), and as we noted, there are strengths of these stimuli (e.g., their applicability to high-stakes situations). However, having access to a large number of experimentally controlled stimuli (e.g., telling truths and lies about both positive and negative statements) featuring targets with diverse characteristics (e.g., race, gender, varying degrees of perceived attractiveness and trustworthiness) will also provide researchers in a number of literatures with a valuable resource. Such a database will allow researchers to focus more on advancing theory and generating

**Table 4** Target-level multiple regression analyses regressing sensitivity, criterion, accuracy, and truth proportion on attractiveness, trustworthiness, and anxiousness

Predictors	Outcome variables
	Sensitivity
	$\beta$
Attractive	.29*
Trustworthy	.04
Anxious	.02
	Criterion
	$\beta$
Attractive	-.07
Trustworthy	.78***
Anxious	-.19**
	Accuracy
	$\beta$
Attractive	.29*
Trustworthy	.01
Anxious	.02
	Truth proportion
	$\beta$
Attractive	-.07
Trustworthy	.78***
Anxious	-.20**

\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$

new findings than on the painstaking process of stimulus set development. Furthermore, having such a large pool of stimuli will aid researchers using more sophisticated analytical strategies, such as signal detection theory, further advancing our knowledge of deception and social perception processes in related fields.

In the present work, we introduced the MU3D, a free resource containing 320 videos of Black and White targets, female and male, telling truths and lies about statements that are positive or negative in valence. The goal of the project was to create a flexible, free resource for researchers interested in conducting work involving well-controlled truth–lie statements presented by targets with considerable diversity (as compared to most existent databases). As such, the MU3D joins other recent database contributions, such as the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015), in providing researchers with high-quality, tightly controlled, well-normed stimuli that also offer target diversity in ways that are often difficult (or even impossible) to achieve in many research environments. We do not expect the MU3D to replace all existing lie detection stimuli. Indeed, the characteristics of the MU3D make it less desirable for research involving criminal or high-stakes deception. However, we believe that the MU3D will create new research opportunities for people who study lie detection, intergroup relations, and social perception more broadly.

Descriptive analyses of the video characteristics (e.g., length) and subjective ratings (e.g., attractiveness) suggest consistency in video length, adequate interrater reliability, and meaningful variability in the subjective ratings. Specifically, video-level analyses indicated that videos with relatively more trustworthy targets were evaluated as being more truthful. Conversely, videos with relatively more anxious targets were evaluated as being less truthful. Target-level analyses echoed this finding, but they also revealed that both sensitivity and accuracy increased for targets greater in attractiveness. Although these analyses were descriptive and exploratory in nature, they enhance the validity of the stimulus set (e.g., anxiousness was negatively correlated with “truth” responding) and suggest numerous avenues for future work. For example, although the videos generated by Black and White target participants did not differ in length, the videos of White targets contained significantly more words than the videos of Black targets. Future research might delve deeper into this finding by exploring whether deception detection videos generated by Blacks and Whites differ in other meaningful ways.

### Advantages of the MU3D

In developing the MU3D, we used the personal description paradigm in order to maximize experimental control and generalizability to everyday deception circumstances. This

method controls for targets’ individual differences (i.e., all targets tell truths and lies) and fully crosses veracity with the valence of the statement. To highlight just one such advantage, we noted earlier the possibility that the valence of a statement could affect both lie detection accuracy and truth bias, and the MU3D enabled us to test this hypothesis. Using the existing database, we conducted independent-samples *t* tests comparing the 160 negative and 160 positive videos on accuracy and the proportion of “truth” responses. The results indicated that positive and negative videos did not differ in terms of accuracy,  $t(318) = -1.28$ ,  $p = .202$ ,  $d = -0.14$ , but that negative statement videos ( $M = .57$ ,  $SD = .18$ ) were less trusted than positive statement videos ( $M = .62$ ,  $SD = .19$ ),  $t(318) = -2.07$ ,  $p = .039$ ,  $d = -0.23$ . Many other questions can be tested with the data provided in the current database (e.g., examining linguistic content among targets with different characteristics), and researchers can further use these stimuli to evaluate other (as yet unmeasured) dimensions of interest in order to explore new questions relevant to their programs of research.

In addition to enhanced experimental control, the MU3D features a relatively sizable number of videos and targets, enabling greater external validity, exploration of target-level characteristics in deception detection, and signal detection analyses. The MU3D will likely be useful both for investigating novel hypotheses (e.g., does valence affect deception detection?) and for revisiting previously demonstrated effects or exploring controversies in the literature. For example, previous research has investigated the effect of target sex on deception detection accuracy, with little consensus (e.g., DePaulo & Tang, 1994; Forrest & Feldman, 2000; Porter, Campbell, Stapleton, & Birt, 2002). Some researchers have reported greater accuracy when judging female as compared to male targets (DePaulo & Tang, 1994; Forrest & Feldman, 2000), whereas other researchers have reported greater accuracy when viewing male as compared to female targets (Porter et al., 2002). Notably, all of the studies mentioned above featured small stimulus sets (DePaulo & Tang, 1994: four female and four male targets; Forrest & Feldman, 2000: eight male and eight female targets; Porter et al., 2002: four male and four female targets) generated by the study authors, which could have varied in numerous, nonsystematic ways. The MU3D can provide additional insight into inconsistent findings, especially those with relatively small stimulus sets. For instance, using the MU3D target-level data, we conducted an independent-samples *t* test comparing accuracy for female and male targets. This analysis yielded a nonsignificant result,  $t(78) = 1.29$ ,  $p = .200$ ,  $d = 0.29$ . Thus, with this relatively large sample of targets, we saw little evidence that target gender affects accuracy. When inconsistent findings like those noted above occur, this usually indicates undiscovered moderators. Discovering such important qualifiers will be imperative for researchers situated at the intersection of gender and deception detection.

Finally, the database includes Black as well as White targets. Because group membership, and race specifically, impacts many facets of intergroup relations and interpersonal sensitivity, including face memory (Hugenberg, Young, Bernstein, & Sacco, 2010), emotion detection (Elfenbein & Ambady, 2002; Kunstman, Tuscherer, Trawalter, & Lloyd, 2016), social interactions (Bergsieker, Shelton, & Richeson, 2010; McConnell & Leibold, 2001), anxiety recognition (Gray, Mendes, & Denny-Brown, 2008), and lie detection (Lloyd et al., 2017), we believe that the possibility of applying the MU3D to literatures beyond lie detection is a notable strength of the database.

### Accessing the MU3D

As we noted above, the MU3D (i.e., video files and associated codebook) can be accessed at <http://hdl.handle.net/2374.MIA/6067>. Before downloading, researchers are required to agree to the terms of use indicated on the website. Upon this agreement, the entire database and associated codebook can be downloaded for free.

**Author note** We thank Britney Crosby, Joia Mitchell-Holman, Kelli Peterman, and Michaela Williams for their help with video collection. Preparation of the manuscript was supported by National Science Foundation Grant BCS-1423765.

### References

- Albrechtsen, J. S., Meissner, C. A., & Susa, K. J. (2009). Can intuition improve deception detection performance?. *Journal of Experimental Social Psychology, 45*, 1052–1055.
- Belot, M., Bhaskar, V., & van de Ven, J. (2010). Promises and cooperation: Evidence from a TV game show. *Journal of Economic Behavior & Organization, 73*, 396–405.
- Bergsieker, H. B., Shelton, J. N., & Richeson, J. A. (2010). To be liked versus respected: Divergent goals of interracial interactions. *Journal of Personality and Social Psychology, 99*, 248–264.
- Bogaard, G., Meijer, E. H., Vrij, A., & Merckelbach, H. (2016). Strong, but wrong: Lay people's and police officers' beliefs about verbal and nonverbal cues to deception. *PLoS ONE, 11*, e0156615. <https://doi.org/10.1371/journal.pone.0156615>
- Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*, 214–234.
- Bond, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin, 134*, 477–492. <https://doi.org/10.1037/0033-2909.134.4.477>
- Bond, C. F., Omar, A., Mahmoud, A., & Bonser, R. N. (1990). Lie detection across cultures. *Journal of Nonverbal Behavior, 14*, 189–204.
- ten Brinke, L., & Porter, S. (2012). Cry me a river: Identifying the behavioral consequences of extremely high-stakes interpersonal deception. *Law and Human Behavior, 36*, 469–477.
- Carton, J. S., Kessler, E. A., & Pape, C. L. (1999). Nonverbal decoding skills and relationship well-being in adults. *Journal of Nonverbal Behavior, 23*, 91–100.
- DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*, 979–995.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74–118. <https://doi.org/10.1037/0033-2909.129.1.74>
- DePaulo, B. M., & Rosenthal, R. (1979). Telling lies. *Journal of Personality and Social Psychology, 37*, 1713–1722.
- DePaulo, B. M., & Tang, J. (1994). Social anxiety and social judgment: The example of detecting deception. *Journal of Research in Personality, 28*, 142–153.
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*, 203–235. <https://doi.org/10.1037/0033-2909.128.2.203>
- Evans, J. R., Michael, S. W., Meissner, C. A., & Brandon, S. E. (2013). Validating a new assessment method for deception detection: Introducing a psychologically based credibility assessment tool. *Journal of Applied Research in Memory and Cognition, 2*, 33–41.
- Forrest, J. A., & Feldman, R. S. (2000). Detecting deception and judge's involvement: Lower task involvement leads to better lie detection. *Personality and Social Psychology Bulletin, 26*, 118–125.
- Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology, 72*, 1429–1439.
- Gray, H. M., Mendes, W. B., & Denny-Brown, C. (2008). An in-group advantage in detecting intergroup anxiety. *Psychological Science, 19*, 1233–1237.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.
- Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization–individuation model: An integrative account of the other-race recognition deficit. *Psychological Review, 117*, 1168–1187. <https://doi.org/10.1037/a0020463>
- Kassin, S. M., Meissner, C. A., & Norwick, R. J. (2005). “I’d know a false confession if I saw one”: A comparative study of college students and police investigators. *Law and Human Behavior, 29*, 211–227.
- Kircher, J. C., Horowitz, S. W., & Raskin, D. C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior, 12*, 79–90.
- Kunstman, J. W., Tuscherer, T., Trawalter, S., & Lloyd, E. P. (2016). What lies beneath? Minority group members’ suspicion of Whites’ egalitarian motivation predicts responses to Whites’ smiles. *Personality and Social Psychology Bulletin, 42*, 1193–1205.
- Leal, S., Vrij, A., Mann, S., & Fisher, R. P. (2010). Detecting true and false opinions: The Devil’s Advocate approach as a lie detection aid. *Acta Psychologica, 134*, 323–329. <https://doi.org/10.1016/j.actpsy.2010.03.005>
- Levine, T. R. (2007). NSF funded cheating tape interviews. East Lansing, MI: Michigan State University.
- Levine, T. R. (2018). Ecological validity and deception detection research design. *Communication Methods and Measures, 12*, 45–54.
- Levine, T. R., Blair, J. P., & Clare, D. D. (2014). Diagnostic utility: Experimental demonstrations and replications of powerful question effects in high-stakes deception detection. *Human Communication Research, 40*, 262–289.
- Levine, T. R., Clare, D. D., Blair, J. P., McCornack, S., Morrison, K., & Park, H. S. (2014). Expertise in deception detection involves actively prompting diagnostic information rather than passive behavioral observation. *Human Communication Research, 40*, 442–462.
- Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the “veracity effect” *Communications Monographs, 66*, 125–144.
- Levine, T. R., Serota, K. B., Shulman, H., Clare, D. D., Park, H. S., Shaw, A. S., ... Lee, J. H. (2011). Sender demeanor: Individual differences

- in sender believability have a powerful impact on deception detection judgments. *Human Communication Research*, 37, 377–403.
- Levine, T. R., Shaw, A., & Shulman, H. C. (2010). Increasing deception detection accuracy with strategic questioning. *Human Communication Research*, 36, 216–231.
- Levine, T. R., Shulman, H. C., Carpenter, C. J., DeAndrea, D. C., & Blair, J. P. (2013). The impact of accusatory, non-accusatory, bait, and false evidence questioning on deception detection. *Communication Research Reports*, 30, 169–174.
- Lloyd, E. P., Hugenberg, K., McConnell, A. R., Kunstman, J. W., & Deska, J. C. (2017). Black and White lies: Race-based biases in deception judgments. *Psychological Science*, 28, 1125–1136.
- Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago Face Database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47, 1122–1135. <https://doi.org/10.3758/s13428-014-0532-5>
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, 98, 185–199. <https://doi.org/10.1037/0033-2909.98.1.185>
- Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, 89, 137–149.
- McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, 37, 435–442.
- Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates ([www.LIWC.net](http://www.LIWC.net)).
- Planalp, S., & Honeycutt, J. M. (1985). Events that increase uncertainty in personal relationships. *Human Communication Research*, 11, 593–604.
- Porter, S., Campbell, M. A., Stapleton, J., & Birt, A. R. (2002). The influence of judge, target, and stimulus characteristics on the accuracy of detecting deceit. *Canadian Journal of Behavioural Science*, 34, 172–185.
- Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: Implications for the recovered memory debate. *Law and Human Behavior*, 23, 517–537. <https://doi.org/10.1023/A:1022344128649>
- Shoda, T. M., & McConnell, A. R. (2013). Interpersonal sensitivity and self-knowledge: Those chronic for trustworthiness are more accurate at detecting it in others. *Journal of Experimental Social Psychology*, 49, 440–443.
- Sporer, S. L. (1997). The less travelled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology*, 11, 373–397.
- ten Brinke, L., Stimson, D., & Carney, D. R. (2014). Some evidence for unconscious lie detection. *Psychological Science*, 25, 1098–1105. <https://doi.org/10.1177/0956797614524421>
- Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: The sequel. A kernel of truth in predictive cheating detection. *Evolution and Human Behavior*, 28, 260–271.
- Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, 24, 239–263.