in a plummeting into depression again, as evidenced in its severity in his eighth-grade year. At present, Drew no longer takes psychotropic medications, as, by his own report, he no longer needs them to manage either anxiety or depression. There is a risk factor that this may no longer be true if he learns that he will not be able to resume using the boys' restrooms and that his attempt at legal redress was not effective.

52. Drew reports that he does not feel comfortable with some of his schoolmates and their more conservative values and attitudes, not trusting the student body as a whole, and perceives that people do not agree with him, even if they don't say it. At the same time, he experiences strong support from other of his fellow students, who would feel fine about him using the boys' bathrooms. He reports one student in particular who came up to him and thanked him for all the work he was doing (related to transgender rights, of which Drew is absolutely dedicated in his work and accomplishments). What can be seen here is a cacophonous experience of his social environment at school—an admixture of rejection and support. By observation, the prohibition of using the boys' bathrooms has tilted the scales towards mistrust and a feeling of lack of safety, which are often responsible for what Drew describes as his "bad" days. If this prohibition was lifted, it could be anticipated that the scale would tilt in the other direction, toward trust and a feeling of acceptance, which has been observed to directly correlate with better psychological functioning in research studies and clinical reports.

53. Another issue to be addressed is Drew's academic performance and experience. Because Drew is a good student, diligent in doing his school work, and demonstrating strong grade performance, it is assumed that the school's policy of limiting Drew to gender-neutral

bathrooms is having no ill effect on his academics. Interviewing Drew about his school performance revealed otherwise. Although he has no records of tardiness in any of his classes, this is because he makes a point of getting to class on time, but may excuse himself midclass to go to the bathroom, which may take anywhere from 10 to 20 minutes away from his class time, depending on classroom distance from the bathroom and weather conditions. Therefore, he might be missing a third of a classroom lecture, and will have to study and complete assignments with these chronic gaps in classroom attendance. He describes what he's missing in one of his classes that involves instructor lectures: "I miss his emphasis on what's important. I don't have that advantage." Other times, he very much has to use the bathroom, but the teacher is saying something very important, so he forces himself to stay. He says this happens about once a month, and cannot be good for his physical health (Note: Although Drew does not report any, many transgender youth end up with chronic urinary tract infections or impacted bowels for this reason). None of Drew's absences will show up as "tardies" in his school records, but the classroom absences remain, and add up to Drew missing approximately a quarter of his classroom learning, a deficit we would not want for any student. Like many transgender students who do not have access to the bathroom that would fit their needs, Drew limits his intake of liquids and tries his best to get through the day without such disruptions, but nature calls, and Drew's choice to use class time for bathroom visits is based on not wanting to call attention to himself when other students are milling about, as would be true during passing times, and not wanting to be late for class, which would necessarily occur in that there is not enough time during passing periods for Drew to get to the bathroom and to the class. Because of these dilemmas, Drew reports that his classroom concentration is compromised—

he is either preoccupied mapping out his bathroom strategies ("I watch the clock instead of paying attention—How long until I can access the bathroom?"), has to attend to uncomfortable physical sensations, or begins to feel frustrated that he has to be in this situation at all and can't just pop in the bathroom between classes like all his other friends. He describes bathrooms as unfortunately becoming a big part of his life, increasing anxiety and lowering concentration, neither of which are good ingredients for learning. It should also be added for consideration that in February, 2017, Drew suffered a knee injury that brought him to the emergency room, with a diagnosis of a knee sprain. This would mean that traveling across campus to use a restroom now became an additional strain, both physically, emotionally, and time wise, which could have been avoided if Drew had had the opportunity during this time to use the boys' restrooms which would have been in closer proximity to his classrooms.

54. Drew expresses anxiety about grades that are not good enough and courses, particularly math and science, which are challenging. Many students feel such anxieties, apart from what bathrooms they are allowed to use, but when the loading for Drew is increased by the intrusion on his concentration and the up ticking in anxiety because of bathroom limits, it gives him less of a leg to stand on in overcoming the worries and working to enhance academic performance, even in difficult subjects. In his own words, "As a kid I was a straight A student. I'm not a straight A student now." He believes his grades would have been higher, not by a lot but by a little in classes where he misses class time because he has to use the bathroom across campus.

55. Regarding the future, Drew reports that if the school's policy is rescinded, "I'd be really happy that they finally came around, knowing that I didn't have to worry. They

would officially be recognizing me as a boy." He reports a 7 out of 10 confidence rating that the present bathroom policy will be rescinded. If, on the other hand, the policy remains in place, he reports it would be hard, he'd feel somewhat depressed, and his hope would diminish. It should be noted that this would most likely not simply be a temporary dip in mood, but could definitely grow into a sustained depression that could very well result in a sense of a foreshortened future and significant drop in motivation, which could be particularly harmful at this cross-section of Drew's life when he is beginning to actively anticipate applying for college and pursuing his aspired career as a medical professional.

56. In conclusion, clinical interviews with Drew indicate a definite pattern of increased psychological stress and mildly impaired academic performance as a result of the instituted bathroom policy of 2015 which prohibited Drew from using the boys' bathroom. While Drew certainly has other sources of anxiety, as any teen would, including stress about college applications, friendships and crushes, and while Drew also has the added minority stress impact of being a transgender youth, it is my professional opinion, based on my clinical expertise and on my observations of Drew, that the implementation of the school's bathroom policy in September 2015 has direct bearing on Drew's levels of anxiety and mild depression. Having also interviewed Drew also about the effects of his parents' divorce and the possibility of direct bullying by peers in his high school experience, both were ruled out as factors causing anxiety or depression, the latter because Drew reports that the divorce is a cordial one to which he has fully adjusted, the latter because Drew reports no direct bullying since his middle school experience. With that said, psychological remedy would most likely be evident if this bathroom policy was revoked and replaced by a policy of allowing all students to use the

bathroom that aligns with their affirmed gender identity. It would not only make for a more comfortable school day for Drew, it would also bolster his confidence as he experienced himself as recognized and supported for the boy he is.

* * * * *

57. I respectfully reserve the right to modify and expand upon my testimony as the facts are developed in this matter.

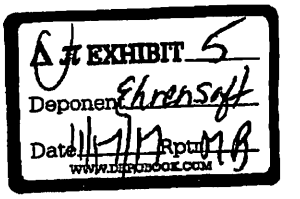
Dated this _21st___ day of September, 2017.

Diane Ehrensaft, Ph.D.

UNITED STATES DISTRICT COURT
FOR THE MIDDLE DISTRICT OF FLORIDA
JACKSONVILLE DIVISION

DREW ADAMS, a minor, by and through his next friend and mother, ERICA ADAMS KASPER, *Plaintiff*, v. THE SCHOOL BOARD OF ST. JOHNS COUNTY, FLORIDA, *Defendants.*	No. 3:17-cv-00739-TJC-JBT

REBUTTAL EXPERT REPORT OF DIANE EHRENSAFT, Ph.D.



1. In preparing this rebuttal report, in addition to the materials listed in the expert report previously submitted on October 2, 2017, I have relied on my review of the expert witness report submitted by Dr. Allan M. Josephson; as well as reviewed the World Professional Association for Transgender Health ("WPATH") Standards of Care, Version 7; the 2017 guidelines for transgender care recently released by the Endocrine Society; the 2015 American Psychological Association guidelines for transgender care; and the 2017 Australian guidelines for transgender care, released in September 2017. In addition, I have reviewed my notes for the interviews I conducted with Drew Adams and a 2017 article authored by Dr. Jack Turban and myself, *Research Review: Gender identity in youth: treatment paradigms and controversies*, which was just published in the Journal of Child Psychology and Psychiatry.

2. The conclusions drawn by the report of Dr. Josephson contain significant methodological flaws and appear to reflect a particular ideology rather than current scientific and medical knowledge regarding gender identity and transgender persons.[1] Those flaws include misuse of statistics, misrepresentation of the studies cited and of the limitations of

[1] The framing and language used by Dr. Josephson in his report is very similar to a position paper entitled "Gender Dysphoria in Children" by the American College of Pediatricians. American College of Pediatricians, Position Statement: Gender Dysphoria in Children (2016), available at, https://www.acpeds.org/the-collegespeaks/position-statements/gender-dysphoria-in-children. The American College of Pediatricians is an association of pediatricians who view being gay or transgender as a disorder, despite the scientific evidence to the contrary. In 2010, Francis S. Collins, M.D., the Director of the National Institute of Health, in a statement made of NIH letterhead, referred to the American College of Pediatricians as a special interest group distorting scientific information to make points against homosexuality, pulling language out of context to "support an ideology that can cause unnecessary anguish and encourage prejudice" John Commins, *NIH Director Raps American College of Pediatricians for Distorting Research on Homosexuality*, HealthLeaders Media (Apr. 16, 2010), available at, http://www.healthleadersmedia.com/physician-leaders/nih-director-raps-american-college-pediatricians-distorting-research-homosexuality.

those studies, and failure to cite studies that disprove or undermine conclusions drawn. This renders the report of Dr. Josephson unscientific and unreliable.

I. Standards of Care and Guidelines Are the Most Reliable Source for Providing Optimal Gender Care in light of Scientific and Clinical Evidence.

3. The basic tenets of care for transgender care and treatment of gender dysphoria are reflected in the WPATH Standards of Care, Version 7; The American Psychological Association Guidelines for Psychological Practice with Transgender and Gender Nonconforming People; the Endocrine Society's Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons: An Endocrine Society Clinical Practice Guideline; and the most recently released standards of care, the Australian and New Zealand Standards of Care for Transgender and Gender Diverse Children and Youth, published in September 2017.

4. Standards of care are constructed specifically to direct practitioners toward best practices in the treatment of their patients, based on existing scientific evidence and professional consensus among entities assigned the task of designing those standards. Practice or clinical guidelines offer recommendations to practitioners to assist them in providing competent care in a particular area of treatment, but are not meant to be as rigorously followed as standards of care.

5. As stated succinctly by the authors of the American Psychological Association Guidelines for Psychological Practice with Transgender and Gender Nonconforming People, "Standards are mandates to which all psychologists must adhere, whereas guidelines are aspirational." (p. 833). However, the construction of both standards of care and guidelines are based on review of available scientific evidence and consensus among the working group of professionals constructing the manuals, a consensus drawn from both review of the scientific

evidence and professional knowledge from clinical practice across senior practitioners in the field.

6. Noteworthy in the area of care for transgender and gender diverse children, adolescents, and adults is the present consistency across the most recent documents released, specifically, the WPATH Standards of Care (2011); the APA Guidelines for Transgender Care (2015), the Endocrine Guidelines for Transgender Care (2017), and the most recent standards of care released, The Australian and New Zealand Standards of Care (September, 2017). Consistent across all these documents is that:

a. Being transgender is a healthy and natural component of the human condition, not a disease;

b. Attempts to alter an individual's gender to fit social expectations are harmful and should not be practiced;

c. Psychiatric co-occurring conditions, prevalent in the transgender population, are typically a result of minority distress and environmentally induced stigma, rather than internal mental disturbance;

d. Positive acceptance, support, and provision of gender-affirming treatments that respect an individual's expressed gender identity promotes healthy physical and psychological outcomes, while lack of acceptance/support and denial of such treatments puts an individual at risk for negative physical and psychological outcomes; and

e. Practitioners should develop a knowledge base and be trained in these gender-affirming practices if they are to be interfacing with gender diverse and transgender patients.

7. Failure to follow these guidelines or standards of care, in a professional community that relies on both to keep them abreast of the most recent scientific discoveries in their field, inform them of best practices, and direct them toward competent care, is typically assessed as substandard practice. This is especially true in situations when existent standards or guidelines from the dominant professional organizations in their field are consistent with each other and in agreement about best practices, as they are regarding transgender care.

8. It should also be noted that within these documents are also references to best practices not just for mental health and health professionals, but for families, schools, religious institutions, and community organizations.

9. Representative of this consistency across major health organizations internationally, itemized here are the major tenets of standards of care, as articulated in the most recently published document: Telfer, M.M., Tollit, M.A., Pace, C.C., & Panga, K.C. *Australian Standards of Care and Treatment Guidelines for Trans and Gender Diverse Children and Adolescents*, Melbourne: The Royal Children's Hospital, 2017. These standards were established on the basis of available scientific and empirical evidence and clinical consensus:

a. "In the past, psychological practices attempting to change a person's gender identity to be more aligned with their sex assigned at birth were used. Such practices, typically known as conversion or reparative therapies, lack efficacy,

and are considered unethical and may cause lasting damage to a child or adolescent's social and emotional health and well-being." (p. 5);

b. "Being trans or gender diverse is now largely viewed as part of the natural spectrum of human diversity." (p 2); and

c. "Increasing evidence demonstrates that with supportive, gender affirming care during childhood and adolescence, harms can be ameliorated and mental health and wellbeing outcomes can be significantly improved." (p. 2).

10. In summary, the role of the mental health professional is to do a thorough assessment and provide avenues for a child or adolescent to explore and consolidate their affirmed gender identity, with additional services offered to parents to strengthen their levels of support to the child, and counsel the youth and parents, in adolescence, about possible available medical interventions.

11. Practices that encourage parents to set arbitrary or inappropriate limits on their children's authentic gender expression or the categorization of a child's persistent declarations of a cross-gender identity as a psychiatric disturbance violate the standards of care. Included in those violations would be reference to the child needing to be met with firm, empathic limits, and redirection and likening the child's gender articulations to childhood insistence on countless things that are not healthy or good for them, as Dr. Josephson suggests in paragraph 24 of his report.

12. It also violates those standards of care to treat a desire to live in accordance to one's affirmed gender identity as an avoidance of challenging developmental hurdles rather than "dealing with struggles on the road to health" with meaningful psychotherapy as an

empathic combination of support/affirmation and encouragement to change and improve," as Josephson suggests in paragraph 32. Although Josephson does not specifically define in that paragraph what he is referring to when he speaks of change and improvement, if change and improvement involve recognizing the "delusion of transgender ideation," as stated in other sections of Dr. Josephson's report, that would be antithetical to the extant standards of care and clinical guidelines of all major health organizations, which clearly state that being transgender is not a disease.

13. The assertion that the current available medical interventions for treating transgender adolescents constitute "eliminating puberty" shows a lack of understanding of the standards of care and medical protocols. (See Josephson report, paragraph 27). There is no existing practice or scientific evidence that puberty could ever be eliminated, as stated in Dr. Josephson's report (See Josephson report, paragraph 27). Instead, best practices are to make available to a transgender youth through medical interventions (puberty blockers and masculinizing or feminizing hormones) the possibility of a puberty more in alignment with their affirmed gender than with the sex assigned to them at birth.

II. Extant Standards of Care and Clinical or Practice Guidelines for Transgender Youth Consistently Endorse a Gender Affirmative Model of Care.

14. The gender affirmative model of care is defined as a model of care recognizing that gender is a combination of biology, environment, and culture and that goals of treatment should be to facilitate a process for a child or youth to live in their legitimate affirmed gender. As stated above, in all these documents gender variations are perceived as a healthy variation among human beings and it is understood that psychological symptoms are most likely a result

of minority stress (i.e., the psychological distress or angst resulting from negative behavior and discrimination targeted at the individual from the social world) rather than disease.

15. Although gender is understood to be a complex interplay of nature, nurture, and culture, it is recognized that for transgender people there is a strong biological underpinning. Gender identity is an internal core component of one's identity, one that may or may not match the sex assigned at birth, and one that is not enforced or legislated by others but internally driven.

16. Care is individualized, with no single form of treatment for all people. For example, in the WPATH Standards of Care, Version 7 (the latest edition), it is no longer required that a person have a "real life" experience (living in the gender role that matches their affirmed gender identity) before receiving medical treatments, and ongoing psychotherapy is no longer a prerequisite to receiving medical care or making a social transition. What has taken the place of either mandatory real life experience or psychotherapy is an interdisciplinary model in which careful assessment and facilitation occurs as the team, consisting of medical and mental health professionals, with the help of the child and family, assesses and acquires knowledge of a child's authentic gender.

III. Legal Statutes Exist Supporting the Present Standards of Care and Practice Guidelines.

17. In nine states (California, Rhode Island, New Jersey, Oregon, Nevada, New Mexico, Illinois, Vermont, Connecticut) and the District of Columbia, legislative statutes exist prohibiting psychotherapeutic practices that attempt to change a minor's sexual orientation or gender behaviors. Similar bills have been introduced in 20 other states, and the Canadian province of Ontario also has legislation banning such clinical practices.

18. In addition to existing laws in a number of states prohibiting discrimination based on gender identity or transgender status in public accommodations, the state of California has also passed legislation that states that every student in the public school system shall be able to use all facilities and engage in all school activities in conformance with their affirmed gender identity, rather than the sex indicated on their birth certificate. Since the passing of that bill there has been no reported instance of any student's privacy being violated by bathroom use according to one's affirmed gender rather than sex assigned at birth.

19. Legislative actions are moving in alignment with the present standards of care and clinical guidelines in assuring the health and well-being of gender diverse and transgender students and prohibiting practices that are implicitly or explicitly advocated in Dr. Josephson's report.

IV. All Clinical Practice Should Involve Careful and Thoughtful Exploration, Rather Than Cursory Endorsement, of a Youth's Initial Reporting About Their Gender.

20. I agree with Dr. Josephson that clinical practice should involve careful and thoughtful exploration rather than cursory endorsement of a youth's initial reporting about their gender (See Josephson, Paragraph 34). Problematic, however, is his assumption that the gender affirmative model of care fails to engage in such practices. The model of care promoted in both the extant standards and guidelines involves careful investigation and exploration of a youth's gender, along with consideration of co-existing psychological issues for a youth that may interface with their gender explorations or self-understandings. No cursory endorsement is involved.

21. The best indication of this model in practice is a consideration of the plaintiff in this complaint, Drew Adams. Having had the opportunity to interview Drew directly and to also review his clinical records, there is sufficient documentation and clinical evidence, along with my own observations, that Drew, with the aid of several mental health and health professionals, has spent much time exploring and bringing into focus his thoughts, feelings, and stresses related to his gender, and with the help of extensive professional care and support came to the realization that his authentic gender is male. In accordance with operationalizing that realization through a social and medical transition, again with continued support from trained professionals, Drew is now only asking that he be allowed to live as the boy he is in every aspect of daily life, which would include access to bathrooms that match his gender, not the sex assigned to him at birth.

V. The Conclusions of Dr. Josephson's Report are Methodologically Unsound.

22. Dr. Josephson relies on incomplete, outdated, and methodologically flawed data, as will be described below, and then extrapolates from that unreliable data to support the view that treatment of transgender children should seek to alter the child's gender identity to conform to the child's sex assigned at birth. That view has no support in the scientific literature or in current medical knowledge and practice, which recognizes that such treatments are harmful and unethical. *See* U.S. Dep't of Health and Human Servs., Substance Abuse and Mental Health Servs. Admin., *Ending Conversion Therapy: Supporting and Affirming LGBTQ Youth* (2015), available at, http://store.samhsa.gov/shin/content/SMA15-4928/SMA15-4928.pdf; American Psychological Association, *Report of the American Psychological Association Task Force on Appropriate Therapeutic Responses to Sexual Orientation* (2009),

available at, https://www.apa.org/pi/lgbt/resources/therapeutic-response.pdf; World Prof. Association for Transgender Health, *Standards of Care for the Health of Transsexual, Transgender, and Gender Nonconforming People* (2011), available at, https://s3.amazonaws.com/amo_hub_content/Association140/files/Standards%20of%20Care%20V7%20-%202011%20WPATH%20(2)(1).pdf.

23. The report misrepresents research relating to the desistence rates among children diagnosed with gender dysphoria. First, Dr. Josephson fails to point out a critical limitation in those studies, which is that those studies focused on children with gender dysphoria (or its predecessor, gender identity disorder), but not transgender youth. Although all transgender youth meet the criteria for gender dysphoria, not all youth diagnosed with gender dysphoria are transgender. Further, Dr. Josephson draws conclusions about transgender children from a sample of children diagnosed with gender dysphoria, which is not the same as it includes children who are not transgender, and fails to recognize that not all transgender children will be captured by a diagnosis of gender dysphoria.

24. Second, a number of key articles that Dr. Josephson relies on in his discussion of the desistence of gender dysphoria have additional methodological weaknesses. For example, in "Psychosexual Outcome of Gender-Dysphoric Children," by Madeleine Wallien and Peggy Cohen-Kettenis, the study started with a cohort of seventy-seven children who had been diagnosed with gender identity disorder, which is now referred to as gender dysphoria. Of that cohort, twenty-three were lost to follow up and for another ten the follow up was conducted with a parent, not the youth. Instead of excluding those children from the statistical analysis, which is a necessary methodical requirement in scientific research, the authors

continued to count them as subjects in the longitudinal study and combined them with those deemed to have "desisted" (i.e., no longer met the diagnostic criteria for gender identity disorder) – resulting in an artificially depressed 27% "persistence" rate. A similar methodological error was made in "Desisting and Persisting Gender Dysphoria After Childhood: A Qualitative Follow-up Study," by Thomas Steensma, et al. That study started with a cohort of fifty-three adolescents who had been diagnosed with gender identity disorder. Of that cohort, twenty-four were lost to follow up. The authors noted in the article that "[a]s the Amsterdam Gender Identity Clinic for children and adolescents is the only one in the country, we assumed that their gender dysphoric feelings had desisted." This causal assumption is clearly flawed, as these adolescents might have many reasons for not returning to the clinic beyond whether they continued to be gender dysphoric, and furthermore, as mentioned above, it is not allowable to count individuals who have dropped out of a study as subjects once lost to the examiner. Further, the critical variables to be measured to determine transgender status, which included measures of gender identity and measures of gender expression, were not the independent variables used in the studies of desisters and persisters, as they should have been if the focus of the study is to determine whether one is transgender or not. Qualifying for a diagnosis of gender identity disorder, the independent variable used in these studies and for which the diagnostic criteria were different than diagnosis of gender dysphoria, fails to meet the standard of measurements necessary to determine transgender status, which includes measures of gender identity and gender expression. Because of those serious flaws, these articles provide no reliable information about the desistance rates for children diagnosed early in life with gender dysphoria.

25. Third, the impetus behind undertaking scientific studies on desistence was to hone the diagnostic criteria used by professionals to more accurately distinguish between transgender youth and youth who are gender-nonconforming. As reflected in the current medical consensus of experts in this field, that goal has been largely achieved. As discussed in "Factors Associated with Desistence and Persistence of Childhood Gender Dysphoria: A Qualitative Follow-Up Study," by Thomas Steensma, et al., four commonly used hallmarks have been identified to differentiate children who are grappling with their preferred gender expressions but not their gender identity from transgender children: (i) the intensity of gender dysphoria; (ii) that the child indicates they are the "other" sex as opposed to wishing to be the "other" sex; (iii) evidence of a significant degree of discomfort with their genitals (body dysphoria); and (iv) age of referral. Dr. Josephson fails to acknowledge the investigators' review of their own research, information which is widely accepted and relied upon by experts in treating transgender children (e.g, cf. D. Ehrensaft, *Gender Born, Gender Made & The Gender Creative Child*).

26. As a result, there is no support for the conclusion that affirming a transgender child's gender identity will cause a child whose gender dysphoria would have otherwise desisted to persist. All data point to the fact that children who underwent an early social transition had already exhibited the objective hallmarks previously mentioned, i.e., were already clearly transgender in their own understanding of themselves and as observed by others, including mental health and medical professionals. Thus, consistent with the standards of care, social transition was the appropriate treatment and supporting those children through a social transition contributed to their overall positive mental health.

27. Lastly, the persister/desister research which Dr. Josephson relies on in his report does not pertain to transgender youth who do not surface with either gender dysphoria or a transgender knowledge of self until adolescence, often triggered by the onset of puberty which feels discordant to the youth.

VI. Transgender Youth Are a Small Percentage of the Population, But That Does Not Render Them Abnormal.

28. Although transgender people are a small percentage of the overall population, Dr. Josephson inappropriately extrapolates that statistic to support the belief that being transgender is not normal and is a disease that must be cured. *See* Josephson Report, para. 22. There are many human variations that are rare or affect only small populations and that are not equated with disease, such as people with high IQs. The rarity of a particular occurrence or trait is just that, evidence of its rate of occurrence within a population; that statistic indicates nothing about whether the occurrence or trait is maladaptive.

29. Stated differently, minority status does not equate with psychiatric abnormality. Presently it is estimated that somewhere between 1 and 2% of the population is transgender, and it is assessed that these are underestimates, due to the reluctance of many to report their transgender status. In addition to the analogy of the comparatively rare number of individuals rated as having superior intelligence, we can also refer to the analogy of handedness. Left-handed people represent only 10% of the population; therefore, individuals who hold this status qualify as a minority population, possessing a variation in brain make-up, but not abnormality.

30. As discussed presently in this statement, and in my prior declaration in this matter, scientific studies and clinical experience demonstrate that being transgender is a normal part of human variation.

VII. Transgender Status Is Not a Mental Disorder.

31. By all existent mental health diagnostic measures, being transgender does not qualify as a mental disease or a delusion, although specifically stated in Paragraph 16 of Dr. Josephson's report and suggested, in referring to transgender ideation as delusion, in Paragraph 43 of his report.

32. In 2013, the DSM-V replaced the DSM-IV. The previous child and adolescent gender diagnosis "Gender Identity Disorder" was removed from the DSM manual and replaced by the "Gender Dysphoria" diagnosis, a diagnostic category that replaces the concept of disorder with the acknowledgement of the stress or distress that may accompany a youth's realization that the gender they experience themselves as being to be discordant with the gender that would match the sex assigned to them at birth.

33. In preparing for the new ICD 11, there has been extensive field study investigation as to whether a childhood gender diagnosis should exist at all, and if it does, whether it should be renamed "gender incongruence" and be removed from chapters on mental disorders, for the precise reason that it is not a disorder in itself.

34. Presently, within the community of health care community there is much debate as to whether a childhood gender diagnosis should exist. While the overwhelming consensus is that gender nonconformity is not pathological, nonetheless some want to retain the diagnosis for practical reasons related to access to care. Specific concerns among those opposing a childhood mental health gender diagnosis are 1) that having such a diagnosis is in tension with the most recent standards of care which consistently de-pathologize gender nonconformity and transgender identity, and 2) that the diagnosis will be misused by those who are ideologically

opposed to the concept of gender diversity and will use the diagnosis to maintain a stance that transgender ideation or identity is a mental illness, promoting an obsolete notion that such experiences represent a mental illness. The latter concern among opponents to the childhood diagnosis is well-founded, as evidenced in the narrative of Dr. Josephson's expert witness report, as when he states for example, "A transgender individual meets the technical, psychiatric criterion for maintaining a delusion: a false, fixed belief, minimally responsive to reason . . . . A deluded person has the freedom to choose beliefs, and should be respected in that choosing, but he/she does not have the freedom to redefine reality" (Josephson report, para. 25). Even those who have been in favor of retaining the diagnosis indicate that the purpose of doing so is not to label a child as disordered but to clinically identify those children and youth who are suffering from stress or distress related to their gender in order to get them the needed treatment for their angst and the supports to live life more authentically in accordance with their experienced gender identity or expressions rather than in accordance to the sex assigned to them at birth.

35. Perhaps of most significance is that no major health organization, including The American Psychological Association, The World Professional Organization for Transgender Health, and the Endocrine Society, presently recognizes transgender identity as a disorder to be cured but rather as a core component of one's identity that may be accompanied by stress or suffering as a result of poor environmental provisions, such as lack of support, respect, or acceptance of the individuals' authentic gender.

36. Scientific evidence that transgender children function within normal range psychologically can be found in the peer-reviewed studies of Dr. Kristina Olson and her

colleagues at University of Washington. Findings were that children who have been identified as transgender and allowed to socially transition to their affirmed gender, when matched with non-transgender peers, showed no differences in psychological functioning from their non-transgender peers, except for a slight elevation in anxiety symptoms, but even then with no areas of psychiatric measures within a clinical range, meaning that the measures indicate these children are within normal range of all psychological areas of functioning measured and indicated rates similar to their non-transgender peers.

37. Nowhere in the standards of care or clinical and practice guidelines is transgender status referred to as a delusion.

* * * * *

38. In conclusion, Dr. Josephson's underlying assumption that being transgender is a disease rather than a natural and healthy variation of humanity is both a violation of present standards of care, in contradiction to both scientific research and clinical or practice guidelines, and a critical flaw in the arguments made in his expert report.

Dated this 2nd day of November, 2017.

Diane Ehrensaft, Ph.D.

Rebuttal Expert Report of Diane Ehrensaft, Ph.D.

Exhibit A – Bibliography

BIBLIOGRAPHY

American College of Pediatricians. "Gender Dysphoria in Children." *Position Statements of the College*, June 2017, https://www.acpeds.org/the-college-speaks/position-statements/gender-dysphoria-in-children.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Washington, D.C.: American Psychiatric Publishing.

American Psychological Association. (2009). *Report of the American Psychological Association Task Force on Appropriate Therapeutic Responses to Sexual Orientation*. Washington, D.C.: https://www.apa.org/pi/lgbt/resources/therapeutic-response.pdf.

American Psychological Association. (2015). Guidelines for Psychological Practice with Transgender and Gender Nonconforming People. *Am. Psychologist* 70:832-864.

John Commins "*NIH Director Raps American College of Pediatricians for Distorting Research on Homosexuality*," *HealthLeaders Media*, April 16, 2010, http://www.healthleadersmedia.com/physician-leaders/nih-director-raps-american-college-pediatricians-distorting-research-homosexuality.

Ehrensaft, D. (2011). *Gender Born, Gender Made: Raising Healthy Gender-nonconforming Children*. New York: The Experiment.

Hembree, W., et al. (2017). Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons: An Endocrine Society Clinical Practice Guideline. *J. of Endocrinology & Metabolism* 102(11):1–35.

James, S., et al. (2016). *The Report of the 2015 U.S. Transgender Survey*. Washington, D.C.: National Center for Transgender Equality.

Steensma, T., et al. (2013). Factors Associated With Desistance and Persistence of Childhood Gender Dysphoria: A Quantitative Follow-Up. *J. of the Am. Acad. of Child & Adolescent Psychiatry* 52:582-590.

Steensma, T., et al. (2011). Desisting and Persisting Gender Dysphoria After Childhood: A Qualitative Follow-up Study. *Clinical Child Psychol. and Psychiatry* 16:499-516.

Substance Abuse and Mental Health Services Administration. (2015). *Ending Conversion Therapy: Supporting and Affirming LGBTQ Youth*. Rockville, MD: U.S. Department of Health and Human Services, http://store.samhsa.gov/shin/content/SMA15-4928/SMA15-4928.pdf.

Telfer, M.M., et al. (2017). *Australian Standards of Care and Treatment Guidelines for Trans and Gender Diverse Children and Adolescents*, Melbourne: The Royal Children's Hospital.

Turban, J., and Diane Ehrensaft. (2017). Research Review: Gender Identity in Youth: Treatment Paradigms and Controversies. *J. of Child Psych. & Psychiatry* doi:10.1111/jcpp.12833.

Wallien, M., and Peggy T. Cohen-Kettenis. (2008). Psychosexual Outcome of Gender-Dysphoric Children. *J. of the Am. Acad. of Child & Adolescent Psychiatry* 47:1413-1423.

World Professional Association for Transgender Health. (2011). *Standards of Care for the Health of Transsexual, Transgender, and Gender Nonconforming People*. https://s3.amazonaws.com/amo_hub_content/Association140/files/Standards%20of%20Care%20V7%20-%202011%20WPATH%20(2)(1).pdf.

CERTIFICATE OF SERVICE

I hereby certify that on November 3, 2017, I caused a true and complete copy of the foregoing REBUTTAL EXPERT REPORT OF DIANE EHRENSAFT, Ph.D., to be served upon the following parties hereto via email:

Robert J. Sniffen (rsniffen@sniffenlaw.com)
Michael P. Spellman (mspellman@sniffenlaw.com)
Terry J. Harmon (tharmon@sniffenlaw.com)
Kevin C. Kostelnik (kkostelnik@sniffenlaw.com)
SNIFFEN & SPELLMAN, P.A.
123 North Monroe Street
Tallahassee, FL 32301
Attorneys for Defendant, The School Board of St. Johns County, Florida

/s/ Tara L. Borelli
Tara L. Borelli (*admitted pro hac vice*)
LAMBDA LEGAL DEFENSE AND
EDUCATION FUND, INC.
730 Peachtree Street NE, Suite 640
Atlanta, GA 30308-1210
Tel.: 404-897-1880 | Fax: 404-897-1884
tborelli@lambdalegal.org

Dovepress
Adolescent Health, Medicine and Therapeutics
Dove Medical Press | This Article | Subscribe | Submit a Manuscript | Search | Follow

Adolesc Health Med Ther. 2017; 8: 57–67.
Published online 2017 May 25. doi: 10.2147/AHMT.S110859
PMCID: PMC5448699

Gender nonconforming youth: current perspectives

Diane Ehrensaft

Department of Pediatrics, University of California, San Francisco, CA, USA
Correspondence: Diane Ehrensaft, Department of Pediatrics, University of California, 445 Bellevue Avenue Suite 302, Oakland, CA 94610, USA, Email dehrensaft@earthlink.net

Abstract

Go to:

Beginning with a case vignette, a discussion follows of the reformulation of theories of gender development taking into consideration the recent upsurge of gender nonconforming and transgender youth presenting for gender services and also in the culture at large. The three predominant models of pediatric gender care are reviewed and critiqued, along with a presentation of the recently developed interdisciplinary model of gender care optimal in the treatment of gender nonconforming youth seeking either puberty blockers or cross-sex hormones.

Keywords: gender nonconforming, transgender, pediatric gender care, puberty blockers, cross-sex hormones

Introduction

Go to:

The field of interdisciplinary treatment for gender nonconforming children and youth has not just expanded at an astronomically fast rate; to switch metaphors, it has rather been such as a tsunami, with a swell of children and families seeking support and services and stretching existing gender clinics and programs at their seams. This cohort of young people includes those who do not accept the sex assignment given to them at birth, those who do not accept their culture's expectations and rules about gender roles and gender behaviors, and those who present with a combination of both.

The case of Daniel is presented to launch this review of current perspectives on gender nonconforming youth. Daniel was 19 years old and in his first year of college (note: all identifying information has been changed to preserve confidentiality. In addition, the patient in the case vignette has provided written informed consent for the publication of the anonymized case details). Just a few months earlier he had announced to each of his parents, who were divorced, that he was transgender. For some years before that, he had been living as a girl, assuming that he was either a "butch dyke" or a masculine identified bisexual young woman. His father and stepmother's response was, "Yes, of course, it makes perfect sense. We'll support you in whatever you need". His mother's response was quite different, "God gave you a body, why would you want to go against God's will? I am so ashamed. What will I ever tell my family? I've always supported you, but I can't do this".

Taking a history, Daniel reported that by the end of his sophomore year in high school he discovered that he was transgender. Before that, he never had the language for who he was. Up until second grade, he, then she with the name Daisy, truly believed that when she reached puberty she would simply switch gears, grow a penis, get a beard, and become a man. From early childhood she dressed like a boy, insisted on wearing her hair short, and was perceived by all as the neighborhood tomboy. When she learned about the physical changes that accompanied female–menstruating, growing breasts, she responded, by her own report: "Whew, I'm so glad I'll never have to go through that". When an older youth disabused her of her misconception, informing her that she would receive no exemption and she would never grow to be a man because she was born a girl, she was temporarily devastated, coming to the realization that she was now doomed to walk the plank of female development. For her, this was a horrible thought. When she actually got her period in the sixth grade, she experienced, with trepidation, that her fate had been sealed – "I'm cooked, there's no turning back now".

In middle school, Daisy had her first girlfriend; she confided in her older brother about her new romance, and he promptly issued her a label, "You're a dyke". Except Daisy kept protesting, "I like boys, too". For high school, Daisy chose to go to a boarding school, the prime reason being that she was tired of going back and forth between two houses in her postdivorce family, and just wanted one place to settle into. It was a Catholic all-girls school and she got in trouble for having a romantic relationship with another girl at school. She persisted in dating girls, just not ones from her school, and through her peer connections first learned about the concept of transgender. She surfed the internet, joined chat rooms, and came to discover that "transgender would be me". Her then girlfriend, beginning to recognize who her partner really was, began referring to Daisy as D. and using male pronouns for D. D. never felt happier. But D. kept it a secret for 2 years, waiting out the end of high school and the opportunity to start a new life in college before affirming a male identity publicly. D. chose a liberal arts college far away from home and within weeks came out at school as Daniel. By Thanksgiving break, Daniel was ready to disclose to his parents, and that circles back to the beginning of the story.

After disclosing to his parents, Daniel then wanted hormones to align his body with his male identity, envisioning surgeries, including top and genital surgery, in his future, but not right then. Daniel's story is presented as an opener to highlight the two questions, "What is your gender?" and "What is to be done once discovered?" that underlie all existent adolescent gender care.

Daniel's case is not a unique one. One might even say that it is emblematic of the increasing number of youth who are seeking professional services, along with their parents, to sort out their authentic gender and discover ways to affirm that authenticity. In most Western cultures gender has historically been considered bedrock: one is assigned a sex at birth, either male or female, typically based on external appearance of genitalia, and this assignment determines one's gender for the duration of that individual's life. Upon entrance into the 21st century, that paradigm of gender bedrock has been hit with a sledge hammer; in its stead, we now have gender as moving boulders, with a sensibility of gender not coming in two boxes, but in infinite varieties, and not necessarily stable over the course of one's lifetime. As this has occurred, providers struggle to keep up with newly emerging theories of gender development and standards of care for the proper care of these youth. Just as an example, the World Professional Association for Transgender Health 7th Edition of the Standards of Care,[1] released in 2011, is already outdated and in the process of being revamped, with the section on children and adolescents in particular need of an update. The needed changes come most significantly in the area of social gender transitions for prepubertal youth, minimum ages for medical interventions, particularly puberty blockers and cross-sex hormones, but also surgeries for individuals before reaching the age of majority. Regarding numbers, the cohort of gender nonconforming youth seems to have expanded exponentially in the most recent decade, as reported by gender programs serving these children throughout North America and beyond.[2,3] In negotiating these phenomenal changes in the gender

terrain, four major areas have needed to be addressed: the necessity of relearning gender so that health professionals can retool themselves to best serve this group of youth; the tensions between the three models of care; the importance of interdisciplinary collaboration in care; the introduction of medical interventions in the care of the youth.

Reformulate theories of gender development in light of gender nonconforming youth

Go to:

Most professionals in the field of gender care have had to unlearn everything taught in training about gender and relearn a new model of gender development. To review the traditional model, children at birth are assigned a sex, male or female, typically based on appearance of external genitalia. If the genitalia were ambiguous in appearance, genital surgical procedures to establish a stable singular sex assignment with matching gender were to be performed as soon as possible, and no later than 18 months. The reasoning behind this, as propounded by Dr John Money and his associates,[4] was that after 18–24 months a child is firm in a core gender identity – I am male, I am female, and thereafter it becomes very difficult to change that identity as it is already cognitively fixed. Once knowing one's gender label, which is both facilitated and mediated by parents' conscious and unconscious messages and reflections, a child's next developmental task is to learn how to "do" gender. Known as gender role socialization, this process is done in close relationship to one's mother and father, with the underlying assumption that all children will have both.[5,6] Within the psychoanalytic paradigm, during this same period a tumultuous drama unfolds, the Oedipal phase – children have intense erotic fantasies about their parents: boys will want to marry their mothers, girls their fathers. Through successful negotiation of these fantasies, facilitated by parents' empathy and boundary setting, children will emerge from the Oedipal phase relinquishing those infantile incestuous desires, firming their own heterosexual identities as they forestall gratification and await an opposite sex partner of their own when they reach adulthood.[7] Within that process they will establish a firm gender identity with a new understanding that one is and always will remain the sex listed on one's birth certificate or assigned early in life (for intersex children).[8] Throughout middle childhood youth will continue to internalize the gender norms of their culture, and learn to conform to them. With the advent of puberty and the entrance into adolescence, a new phase of gender consolidation occurs as youth awaken to their adult sexual urges and prepare for their gender-divided roles as men or women.

Within the traditional model of gender development, if this developmental trajectory takes a course other than that described above, there is cause for concern for the child, along with scrutiny of the parents, as parents are held accountable for the child's anomalies. To quote Robert Stoller, a pioneer in the treatment of gender disorders in youth in the 20th century,[9] speaking of "primary transsexual" boys (those nonintersex boys who have been feminine from the first year of life): "As an infant, such a boy usually has an excessively intimate, blissful, skin-to-skin closeness with his mother. This, unfortunately, is not interrupted by his father, a passive distant man who plays no significant part in bringing up his son" (p. 16). In family situations like the one inscribed above by Stoller, professional help was recommended to cure the youth's gender anomalies and to treat the parents so they cease veering their child's gender development in wrong directions because of their own internal conflicts.

For a theory of development to be robust, it should be evident in empirical observation or investigation. The traditional theory of gender development and disordered gender, which is still in use by many, fails that test, for the following reasons:[10]

- Many individuals continue renegotiating their gender throughout childhood or adulthood, with no observable detriment to their mental health;
- Youth may establish a gender identity in concordance with their assigned sex, be firm in that identity, yet not embrace a heterosexual identity, with no aspersion on their emotional well-

being. Gender development and sexual identity development are two separate developmental tracks, albeit crossing at certain points.

- Whereas core gender identity is typically concordant with assigned sex based on observable external genitalia, for a minority of people this is not the case, with increasing evidence that gender identity lies not between our legs, in our genitalia and primary sex characteristics, but in our brains and minds.[11]
- Therefore, one's assigned sex at birth may differ from one's core gender identity, not because of poor parental handling or infantile confusions, but because of brain and mind gender messages overriding signals from genitalia, chromosomes, or parental expectations. Recently, this phenomenon of mind over matter has been referred to as "neurological sex", defined as a uniform standard of legal sex based on gender identity, in which brain messages are privileged over anatomy and chromosomes in determining an individual's authentic gender.[12]

In contemporary versions of gender development theory that take into account gender variations as a normal part of the human condition, the understanding is that the sex assigned at birth may match the gender a youth will eventually know themselves to be, but it might not. Each child is presented with a developmental task of weaving together threads of nature, nurture, and culture to establish their individual and unique authentic gender self. This self will be composed of both gender identity – who I know myself to be as male, female, or other, and gender expressions – how I choose to perform my gender, including clothing choices, activity preferences, friendship choices, and so forth. Recently, this transactional relationship between nature, nurture, and culture in gender development has been referred to as the gender web,[13] broken up into components that consist of the items in Table 1.

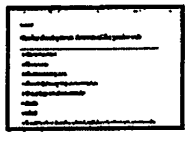


Table 1
Gender development: elements of the gender web

In this contemporary model of gender development, added to the three dimensions of nature, nurture, and culture is the fourth dimension: time. Each child alters their gender web as they weave together nature, nurture, and culture, "over time". In other words, gender is neither fixed by age 6, as in the traditional model, nor static throughout all stages of child and adult development, thus explaining how an individual at age 40 or 50 could come to the realization that the gender they had identified as being is no longer a good fit. It is also recognized that gender development is a discrete and separate track from development of one's sexual identity, and typically proceeds it in a youth's development.

In this model the role of parents and socialization agents is not to shape or reinforce a child's gender identity or expressions, but rather to facilitate it, mirroring back to the child the messages that the child communicates about their preferred gender expressions and articulated gender identity, which may or may not be in concordance with the sex assigned to the child at birth. With the advent of adolescence, it is recognized that some youth's gender trajectories may benefit from medical interventions, including puberty blockers (gonadotropin-releasing hormone [GnRh] agonist) and cross-sex hormones to bring the youth's body in better alignment with their affirmed gender identity.[14] To that end, the model of care that extends from this contemporary theory of gender development is one that strongly relies on interdisciplinary care, especially between mental health and medical providers as they address the holistic medical and psychosocial needs of the emergent cohort of gender nonconforming youth from the perspective of both their psychological and physical development.

Major mental heaith treatment models for gender nonconforming children and youth

Go to:

As of the second decade of the 21st century, three major treatment models are available for addressing the needs of gender nonconforming children and their families, with overlapping premises based on the contemporary model of gender development outlined above but with distinct differences between them. The first model, represented in the work of Drs Susan Bradley and Ken Zucker, assumes that young children have malleable gender brains, so to speak, and that treatment goals can include helping a young child accept the gender that matches the sex assigned to them at birth. The second model, represented in the work of practitioners in the Netherlands, allows that a child may have knowledge of their gender identity at a young age, but should wait until the advent of adolescence before engaging in any full transition from one gender to another. The third model, represented in the work of an international consortium of gender affirmative theoreticians and practitioners, allows that a child of any age may be cognizant of their authentic identity and will benefit from a social transition at any stage of development. To situate and compare each of the three models, a typical referral that may come the way of a gender specialist, regardless of their orientation, is presented, with the assumption that this potential patient may be in need of services from a young age through adolescence:

> *Hi Dr, I came across your information while I was researching for my son.*
>
> *He recently just turned 4 and wants to be a girl and is only drawn to girl toys/clothes for the past 2 years.*
>
> *We have not spoken with a professional doctor. But wanted to reach out early and find ways we as parents can support him.*
>
> *Please let me know if you could help.*
>
> *Thank you!*

Dialing back a generation, if this child's name was Kyle and the same query came to a mental health professional participating in, for example, Dr Richard Green's clinic at the University of California Los Angeles, the treatment recommended and then implemented could very well have looked like this:

> *When he was five, Kyle entered a behavior modification program. [...] Kyle received blue tokens for "desirable" behaviors [...] red ones for "undesirable" behaviors [...]. Blue tokens were redeemable for treats [...]. Red tokens resulted in a loss of blue tokens, periods of isolation, or spanking by father.*[15]

Setting a precedent for other clinicians of the time treating children who presented as gender nonconforming, Kyle's treatment at the UCLA program is emblematic of the model implemented during this era, with the goal of helping children accept the sex assigned to them at birth and adopt the culturally defined appropriate gender behaviors that would match that sex assignment, in alignment with the traditional model of gender development. Underlying the treatment was the intent of warding off a homosexual outcome for young effeminate boys. It should be mentioned that this model is still practiced today, referred to by some as the reparative model.

Focusing now on contemporary approaches that stand in contrast to the above mode, all of which are to be differentiated from the UCLA program, the three major models, outlined earlier, are typically referred to, in order of presentation, as the following:

- The "live in your own skin" model
- The watchful waiting model
- The gender affirmative model

Below is a review of the manner in which each of these models would approach the treatment of a child or youth who is presenting as gender nonconforming, in their gender identity, gender expressions, or both.

The "live in your own skin" model

As mentioned earlier, this model was developed by Drs Susan Bradley and Ken Zucker at the Center for Alcoholism and Mental Health gender clinic in Toronto.16 The treatment goal of facilitating a young child accepting the gender identity matching the sex assigned to that child at birth, based on the supposition that younger children, in contrast to older youth, have a malleable gender brain, is tied to a medical–social rationale. Specifically, being transgender is a harder way to live one's life, both because of social stigma and potential requested hormonal treatments and surgeries to align a youth's body with their transgender identity. Given the perceived plasticity of the young child's gender brain, best practice would be to introduce interventions to help a child accept the sex assigned to them at birth as their gender identity, with no harm done and indeed added benefit to their psychological and social well-being. As explained by Dr Zucker, employing this strategy results in lowering the odds that "as such a kid gets older, he or she will move into adolescence feeling so uncomfortable about their gender identity that they think that it would be better to live as the other gender and require treatment with hormones and sex reassignment surgery".17 In addition to presuming gender identity malleability in young children, the model also assumes that parents' own conflicts or issues about gender likely contribute to a young child's gender dysphoria. With the parents' consent, the "live in your own skin" model employs a combination of behavior modification, ecological interventions, and family system restructuring to facilitate the child arriving at a place of accepting the gender matching their sex assigned at birth. Practices could include taking away cross-gender toys at home and replacing them with "gender-appropriate" toys, altering children's playmate choices to include more same-sex contacts, enrolling the children in "gender-appropriate" activities, encouraging the like-sex parent to become more actively involved and the opposite-sex parent to step back in relationship to the child, and offering psychotherapy to both the child and parents. The aim of treatment of the child is to explore the child's gender and solidify a "live in your own skin" outcome, and the treatment with the parents is aimed at investigating conflicts or psychological issues stemming from or contributing to the child's gender dysphoria. If by the arrival of puberty a child is still exhibiting cross-gender identifications and expressing a cross-gender identity, that child should be supported in transitioning to the affirmed gender, including receiving puberty blockers and hormones, once it is assessed through clinical interviews and psychometric testing that the affirmed gender identity is authentic. The reasoning behind this shift in adolescence is as follows: 1) by adolescence it is too late to intervene in facilitating a child living in their own skin, as the sensitive period of malleable brain development of gender has closed; 2) this individual can now be reliably identified as one of the small minority of youth who persist with a cross-gender identity from early childhood into adolescence, an indicator that this identification will most likely remain stable into adulthood. In the live in your own skin model, the parent reaching out for support of her 4-year-old son might be encouraged to engage in the treatment program outlined above, with the goal of helping her child accept that he is a boy, not a girl and with the intent of warding off a transgender outcome.

The watchful waiting model

The "watchful waiting" model was designed by the members of the interdisciplinary team at the Amsterdam Center of Expertise on Gender Dysphoria, VU University Medical Center, under the leadership of Dr Peggy Cohen-Kettenis. Borrowing from the medical use of GnRH agonists for children exhibiting precocious puberty, the Netherlands team is responsible for introducing the use of puberty blockers for gender purposes, to put a pause on pubertal growth and allow more time for a youth to explore their gender and consolidate their adolescent gender identity, with the future

possibility of cross-sex hormone therapy to align their bodies with their affirmed gender identity. In contrast to the live in your own skin approach, a young child's demonstration of gender nonconformity, be it in identity, expressions, or both, is not to be manipulated in any way, but observed over time. If a child's cross-gender identifications and affirmations are persistent over time, interventions are made available for a child to consolidate a transgender identity, once it is assessed, through therapeutic intervention and psychometric assessment, as in the best interests of the child. These interventions include social transitions (the shift from one gender to another, including possible name change, gender marker change, and gender pronoun changes), puberty blockers, and later hormones and possible gender-affirming surgeries. No attempts are made to alter a child's gender identity or expressions; yet it is postulated in this model that it would be better to hold off until puberty on any social transitions of a child from one gender to another, and instead give them safe spaces to fully express their gender as they prefer before facilitating any full gender transitions.[18,19] The rationale for holding off on any social transitions until adolescence is not to ward off a transgender identity but rather that 1) it would be advantageous that a child experiences the first stages of physical puberty for that child to best make a determination of the gender that feels most authentic to him/her; 2) given developmental stages of childhood, facilitating a social transition from one gender to another at a young age may create a form of cognitive constriction – the child may be prematurely blocked from considering any other possibilities once moved into a cross-gender status and socially constricted from further childhood gender exploration because now they know the cross-gender identity is what everyone has come to expect from them; 3) socially transitioning a child at a very young age may preclude the child from maintaining a realistic understanding of their body and historical status – as a penis-bodied (once a boy) or a vagina-bodied (once a girl) person. In informing their practices, this model, like the live in your own skin model, relies on the data gathered about "persisters" and "desisters", both at their own clinic in the Netherlands and in other international studies, particularly those conducted at the Centre for Addiction and Mental Health (CAMH) gender program in Toronto. In the most recent review of these studies, it was found that 63% of the children seeking services at a gender clinic at a young age, and diagnosed with gender dysphoria, no longer had that diagnosis at puberty, while 37% did have the diagnosis consistently from early childhood to adolescence.[20] Since a large majority of gender nonconforming young children seeking services at gender clinics desist in their gender dysphoria by adolescence, best practices would be to wait and see if the child persists into adolescence before making any significant changes in a child's gender identity.

During the preadolescent waiting period, the children are followed carefully by the clinical team in the watchful waiting model, with the support of outside therapists in the community (which is required before a child can receive medical services), to assure that the children are growing well and getting their emotional needs met, and in preparation for later transitioning and medical interventions if the child proves to be a good candidate. Like in the live in your own skin model, the children going through the program also receive a full battery of psychological tests, documenting not only their gender status but also their cognitive–social–emotional functioning. Some of these instruments are delivered to the children directly, some to their parents or teachers.

If the mother asking for help with her 4-year-old were to attend the Amsterdam clinic with her child, the team might do an assessment and advise that the 4-year-old be followed over time, with the understanding that if her son's declarations of wanting to be a girl persisted over time and if he continued to be drawn only to "girl" toys and activities, consideration of puberty blockers to buy more time to explore gender could certainly happen later, but for now it would be best to let her son continue to be a son free to explore whatever activities he enjoyed, with no corrections on his expressed desire to be a girl.

The gender affirmative model

The third model of care, the gender affirmative model, is closely aligned with the watchful waiting model but in opposition to the live in your own skin model. Where the gender affirmative model parts ways with the watchful waiting model is in the waiting part.

The gender affirmative model is defined as a method of therapeutic care that includes allowing children to speak for themselves about their self-experienced gender identity and expressions and providing support for them to evolve into their authentic gender selves, no matter at what age. Interventions include social transition from one gender to another and/or evolving gender nonconforming expressions and presentations, as well as later gender-affirming medical interventions (puberty blockers, cross-sex hormones, surgeries). A particular set of premises informs the model, as listed in Table 2.

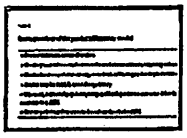


Table 2
Basic premises of the gender affirmative model

The model is informed by the contemporary theory of gender development outlined above, with a recognition that although gender evolves over the course of a lifetime, gender identity appears to be a relatively more stable and consistent construct compared to gender expressions. Gender health is defined as a youth's opportunity to live in the gender that feels most real and/or comfortable, or, alternatively, a youth's ability to express gender with freedom from restriction, aspersion, or rejection.21 When considering a child's gender status, attention is paid to both gender identity and gender expressions, with the understanding that a child's gender identity may communicate something very different about the child than a child's gender expressions might.

Therapeutic goals in the gender affirmative model include:

- Facilitating an authentic gender self
- Alleviating gender stress or distress
- Building gender resilience
- Securing social supports

In contrast to the first two models, no assumption is made that every child exhibiting a gender nonconforming presentation is in need of mental health treatment. Because of the emphasis on social factors affecting the youth, interventions may be targeted at the surrounding environment, rather than the child's individual psyche. This might include interfacing with schools, social and religious institutions, and policy-making bodies to remove the "social" pathology impinging on the child, such as transphobic attitudes and responses, gender policing, or bullying and harassment. Relatedly, parent consultations often take precedence over individual treatment of the child,22–24 with provision of services to help a parent make sense of their child's gender nonconformity, work through any extant conflicts and anxieties about their child's gender, and move toward acceptance of their child.

Individual treatment for the child is indicated for one of five reasons: 1) to assess a child's gender status; 2) to afford the child a "room of their own" to explore their gender; 3) to identify and attend to any co-occurring psychological issues; 4) to address and ameliorate a child's gender stress or distress; 5) to provide sustenance in the face of a nonaccepting or rejecting social milieu, which might include family, school, religious institution, or community. Some professionals working in this model will call on psychometric or projective measures to gather information about the child; others will rely on observation, play, interviewing, and dialog. If assessment instruments are employed, every effort is made to use protocols that do not rely on binary measures of gender (e.g., Are you a boy or a girl?) and

are not pathology oriented, but instead assess strengths as well as weaknesses and differentiate between gender expressions and gender identity.

The basic therapeutic tenet of the gender affirmative model is quite simple: When it comes to knowing a child's gender, it is not for us to tell, but for the children to say. In contrast to the watchful waiting model, once information is gathered to assess a child's gender status, action is taken to allow that child to exercise that gender. Therefore, if after careful consideration, it becomes clear that a young child is affirmed in their gender, demonstrating that the gender they know themselves is different than or opposite to the gender that would match the sex assigned to them at birth, the gender affirmative model supports a social transition to allow that child to fully live in that gender, whether that child is 3, 7, or 17 years old. Such decision-making is governed by stages, rather than ages, both for social transitions and later for medical interventions. Once the child's gender comes into clear focus, which is posited as happening with a child of any age, no need is seen to hold off until adolescence to affirm that gender. This viewpoint is informed by data indicating the psychological harm that can be done, including heightened risk for generalized anxiety, social anxiety, oppositional behaviors, depression, compromised school performance, if a youth experiences themselves living in a gender that is inauthentic to them.25

In the gender affirmative model, the mother of the 4-year-old querying about her son's cross-gender interests would be invited in to the consultation room, along with any other parenting figure involved, to report more about what she had been observing in her child's behaviors from infancy to the present; to determine whether her son is showing any signs of stress or distress about his interest in all things girly things; to explore whether her child is indicating cross-gender expressions vs identity. If there was evidence of stress or distress, by parents' report, or if the parents desired to get a clearer picture of their child's gender status, the family would be invited to bring their son in for observation and play sessions. There would then be the opportunity to reflect, in collaboration with the parents or caregivers, on any evidence that this child was consistent in cross-gender declarations, as in "I'm a girl, not a boy", and that these declarations were persistent over time and not attributable to any other problems in life. If that evidence made clear that this child was communicating about a cross-gender identity rather than desired cross-gender expressions, and if the parents were supportive of their child's gender identity affirmations, it would not be found necessary to recommend to this mother that she wait until puberty to take action regarding her child's gender identity. Instead, a present social transition to the gender that was more authentic for this child, in this case, female, would be considered. If, on the other hand, the child was happy as he was, if given the latitude to play with whatever he wanted and wear whatever he desired, as a boy, the recommendation to the mother might be to give her son the opportunity to express his gender freely, with the opportunity to return for services as requested. Along with this recommendation would be a reminder that all that can be known is the cross section of this child's gender as he presents it at age 4, a gender that may evolve into another configuration later in childhood, at which point a new assessment may be in order.

Critique of the three models

In brief, the live in your own skin model has been challenged as causing potential harm to gender nonconforming youth. A Canadian study conducted by Wallace and Russell assessed that in the living-in-your-own-skin model "there appears to be an enhanced risk of fostering proneness to shame, a shame-based identity and vulnerability to depression."26 Major health organizations, including the World Professional Association for Transgender Health, the American Psychological Association, and the American Psychiatric Association, have issued statements stipulating that mental health professionals are not to engage in practices that attempt to alter the gender expressions or identity of an individual, including children and adolescents. The watchful waiting model is a highly respected model of care worldwide, offering careful and cautious procedures; but it has run into a snag: many

contemporary families in the Netherlands are not content to hold their children back from social transitions until puberty, and have, through both local and international support networks of parents and professionals, proceeded to facilitate their children's social transitions without awaiting clinical approval or waiting until puberty arrives. Parents do this not because they dismiss professional care, but because evidence is accruing that young children thrive when given permission to live in the gender that is most authentic,27,28 and are at risk for symptomatic behaviors if prevented from doing so. At the same time, the watchful waiting model is effective in its thorough attention and assessment of the child over time, integrating the services of mental health and medical professionals.

The gender affirmative model is questioned by some on the basis of the lack of evidence-based data that indicates that young children can reliably communicate and have self-knowledge of a transgender identity or benefit from a social transition. There is also concern that the model of listening to the children puts too much weight on a child's self-report. This is a valid concern, and to address it the self-report is embedded within a collaborative model with the child as subject and the collaborative team including the child, parents, and professionals. Together, the team will be making informed determinations about the most appropriate gender pathways to promote a child's gender health, be it a gender social transition, expanded opportunity to express gender in ways that feel authentic to the child, or deeper exploration of underlying issues that may be presenting as gender stress or distress. Such determinations typically involve extensive consultation and observation, but with no requirement for ongoing psychotherapy or psychometric testing, in comparison to the other two models.

Integration of medical and mental health care in adolescence

Go to:

All of the three models of care referenced earlier share in common the administration of hormonal treatment in adolescence. The first category would be consideration of GnRH agonists (puberty blockers) to put a temporary pause on puberty, providing a youth with additional time to explore gender or, alternatively, warding off an unwanted puberty. The latter is particularly true for youth who socially transitioned early in life, living consistently in their affirmed gender from a young age; in those instances administration of puberty blockers could be considered a form of continuity of care, from social transitions to hormonal intervention. The second category includes feminizing or masculinizing hormones to bring a youth's body in better alignment with their affirmed gender identity. The minimal age for being eligible for such treatments may vary among approaches and indeed among clinics adopting the same approach, but there is common agreement that these treatments are in the best interests of the child who has a documented transgender identity.29 It should be noted that there is probably no other aspect of adolescent care in which the medical and mental health professionals are so vitally interdependent in both assessment and treatment of the youth.30 The reason for this is that each of the interventions has vital interconnected psychological and medical components, requiring an integration of medical evaluation and mental health assessment both to determine appropriateness, assess any medical or psychological impediments to treatment, and monitor follow-up, in terms of effects and supports over time as the youth is administered either the puberty blockers or hormones.

The role of the medical professional is first to assess the youth's level of puberty development, with an assessment of physical readiness for considerations for puberty blockers, which can be administered as soon as the youth enters Tanner Stage 2 of puberty. The medical professional will be responsible for ordering the lab work and bone density scans necessary to monitor a youth's progress and also to screen for any medical counter-indications to administering the blockers. As RnGH agonists are a completely reversible procedure regarding development of secondary sex characteristics, the medical provider will not need to worry about untoward permanent effects in that regard if the youth decides to go off blockers and return to the unfolding of a physical puberty in concordance with the sex assigned at birth. It should be noted, however, that the provider will need to alert the child and family about any side- or long-term effects of RnGH agonists, including effects on bone mineral density and overall

bone health. If, on the other hand, the youth decides to proceed with cross-sex hormones to affirm a gender identity not in concordance with the sex assigned at birth, the medical provider will then be faced with the task of determining if the youth is a good candidate for this next step of treatment. Some youth will have already gone through full puberty before discovering or communicating to others a transgender identity, and the medical provider will be faced with the same task with these youth, with the added feature of explaining to the youth that certain of the developed features of the puberty they have already gone through will not disappear as they go through a second puberty on cross-sex hormones. In either case, cross-sex hormones involve a weightier decision than puberty blockers, as these interventions are only partially reversible in terms of secondary sex characteristics, so the provider will want to be cautious and judicious in determining if cross-sex hormones are appropriate for a particular youth.

This is where the mental health professional enters. In all of the models of gender care, the mental health professional is asked to weigh in as to 1) the authentic gender identity of the youth or level of gender dysphoria exhibited by the youth; 2) the youth's level of maturity and ability to assent to and follow through on the recommended hormonal treatment; 3) the evidence of any coexisting psychological conditions that might interfere with the hormone treatment or that alternatively might bear no weight on the requested treatments or even be alleviated by the hormonal interventions; and 4) the level of family support and willingness to consent to the treatment. In consultation with the medical professional, a decision will be made as to whether a youth is a good candidate for either puberty blockers or cross-sex hormones.

Another critical task for the medical-mental health team is the necessary discussion of fertility implications for each of these interventions. Although advances are being made in reproductive medicine to preserve immature gametes or reproductive tissues for later reproduction, at this point in history a child who begins puberty blockers at Tanner Stage 2 and proceeds directly to cross-sex hormones will be rendered infertile. Administration of testosterone or estrogen to a postpubertal adolescent may compromise a youth's later fertility, or might require going off the hormones for a period of time if a transgender youth who has not had gonad or genital surgeries later in life desires to have a genetically related child. Alternatively, a youth can bank gametes for the future before going on a course of cross-sex hormones, which is a medical possibility but also a psychological challenge for many transgender youth who find this antithetical to their affirmed gender status, requiring a transgender female to attend a fertility clinic and masturbate or a transgender male to undergo a gynecological vaginal ultrasound. Exploring fertility issues before making decisions about blockers or hormones are necessary but sensitive discussions to be had with both the youth and parents, and are best done with the presence of both a medical and a mental health professional who together can provide medical and psychological counsel to the family in this decision affecting later family-building.[31]

Not only is there no other aspect of adolescent care where the teamwork between medical and mental health provider is critical; there is no other domain of youth services in which a mental health provider is so actively involved in medical decision making. Where this has surfaced most recently is in the recent emergence of youth in gender clinics who present as neither male nor female, but rather gender nonbinary or "in the middle", adopting the platform of the multiplicity of gender. The challenge is when these youth ask for a particular medical intervention that achieves that goal of a middle ground – perhaps a touch of testosterone, or chest surgery with no other intervention and a chosen pronoun of "they" rather than "he" or "she". These are new horizons for both medical and mental health professionals today, and there is a mutuality, therefore, in the medical professional training the mental health professional while the mental health professional is in turn training the medical professional in order to integrate the biopsychosocial aspects of care to include the gamut of all the gender nonconforming youth presenting for care.[32]

With that said, it has proved to be critical that mental health professionals involved in this team work be trained gender specialists, with a basic understanding of the medical interventions involved in transgender care, expertise in assessing gender dysphoria and identifying a youth's gender identity, and recognition of psychological issues other than gender that might drive a youth's request for a hormonal treatment. For example, a nurse practitioner on a gender team had administered a puberty blocker implant, Supprelin, which could stay in place for a year, after receiving a letter of support from a trained mental health expert recommending such treatment for this youth who presented as gender dysphoric and in need of further exploration of his gender before going forward with puberty. Over the course of the following year, he failed to return for follow up visits. A year had gone by and it was now time to replace the implant, which the nurse practitioner was prepared to do. The mental health member of the team first did a follow-up evaluation of the youth and discovered that he had made no efforts to explore his gender any further, with his motivation to continue on blockers driven by a desire to remain prepubertal for as long as possible. With the psychologist's guidance, the medical provider was able to recognize that the medical intervention as it stood was inappropriate for this youth. The interdisciplinary team informed the youth that he would be able to receive a new implant only if he was simultaneously working with a mental health gender specialist to further explore his gender identity. If that condition was met, once the twelve additional months on the puberty blockers was completed, the youth would then have to make a determination of which puberty path he would take – cross-sex hormones or the unfolding of his male, testosterone-producing puberty.

Conclusion

Go to:

In the course of only two decades, sophisticated models for the care of gender nonconforming and transgender youth have evolved. There is an urgent need to provide more research data documenting the efficacy of these different programs, but the recent findings of the Amsterdam group provide hope that the care, particularly within the watchful waiting and gender affirmative models, is promoting gender health. In the Dutch authors' words, the treatment, including puberty suppression, cross-sex hormones, and then in adulthood gender affirmation surgery, "leads to improved psychological functioning of transgender adolescents. While enabling them to make important age-appropriate developmental transitions, it contributes to a satisfactory objective and subjective well-being in young adulthood".33 The authors propose that not only early medical intervention, but also a comprehensive multidisciplinary approach contributes to the youth's gender health. Reflecting back on Daniel, the youth introduced at the opening of this review, the ability of professionals to aid youth such as Daniel in getting his authentic gender into focus and providing the appropriate treatments to bring that gender in alignment with his body is the key to overall well-being for all youth seeking professional gender care.

Footnotes

Go to:

Disclosure

The author reports no conflicts of interest in this work.

References

Go to:

1. World Professional Association for Transgender Health . Standards of Care for the Health of Transsexual, Transgender, and Gender Nonconforming People. WPATH; 2011.

2. Hidalgo MA, Ehrensaft D, Tishelman AC, et al. The Gender affirmative model: what we know and what we aim to learn. Hum Dev. 2013;56:285–290.

3. Zucker KJ, Wood H, VanderLaan DP. Models of psycho-pathology in children and adolescents with gender dysphoria. In: Kreukels BPC, Steensma TD, de Vries ALC, editors. Gender Dysphoria and

Disorders of Sex Development: Progress in Care and Knowledge. New York, NY: Springer; 2014. pp. 171–192.

4. Money J, Ehrhardt A. Man & Woman, Boy & Girl: Gender Identity from Conception to Maturity. Northvale, NJ: Jason Aronson; 1996.

5. Diamond M. Sex, gender, and identity over the years: a changing perspective. Child Adolesc Psychiatr Clin N Am. 2004;13(3):591–607. [PubMed]

6. Tyson P. A Developmental line of gender identity, gender role, and choice of love object. J Am Psychoanal Assoc. 1982;30(1):61–86. [PubMed]

7. Freud S. Three Contributions to the Theory of Sex. New York, NY: EP Dutton; 1962.

8. Fast I. Aspects of core gender identity. Psychoanal Dial. 1999;9(5):633–661.

9. Stoller R. Presentations of Gender. New Haven, CT: Yale University Press; 1985.

10. Ehrensaft D. Gender Born, Gender Made. New York, NY: The Experiment Press; 2011.

11. Diamond M. Sex and gender: same or different? Feminism Psychol. 2000;10(1):46–54.

12. Vanderhorst B. Whither lies the self: intersex and transgender individuals and a proposal for brain-based legal sex. Harvard Law & Policy Review. 2015;9:241–274.

13. Ehrensaft D. From gender identity disorder to gender identity creativity: true gender self child therapy. J Homosex. 2012;59(3):337–356. [PubMed]

14. Sherer I, Rosenthal S, Ehrensaft D, Baum J. Child and adolescent gender center: a multidisciplinary collaboration to improve the lives of gender nonconforming children and teens. Pediatr Rev. 2012;33(6):273–276. [PubMed]

15. Green R. The 'Sissy boy' Syndrome and the Development of Homosexuality. New Haven, CT: Yale University Press; 1987. p. 295.

16. Zucker K, Bradley S. Gender Identity Disorder and Psychosexual Problems in Children and Adolescents. New York, NY: The Guilford Press; 1995.

17. Cross JM. Metro News. Toronto: 2015. Mar 18, Outcry prompts CAMH to review its controversial treatment of trans youth. online. Available from: http://metronews.ca/news/toronto/1315743/outcry-prompts-camh-to-review-its-controversial-treatment-of-trans-youth.

18. Cohen-Kettenis PT, Pfäfflin F. Transgenderism and Intersexuality in Childhood and Adolescence. Thousand Oaks, CA: Sage Publications; 2003.

19. Steensma TD, McGuire J, Kreukels BP, Beekman AJ, Cohen-Kettenis PT. Factors associated with desistence and persistence of childhood gender dysphoria: a quantitative follow-up study. J Am Acad Child Adolesc Psychiatry. 2013;52(6):582–590. [PubMed]

20. Steensma TD, Biemond R, de Boer F, Cohen-Kettenis PT. Desisting and persisting gender dysphoria after childhood: a qualitative follow-up study. Clin Child Psychol Psychiatry. 2011;16(4):499–516. [PubMed]

21. Ehrensaft D. The Gender Creative Child. New York, NY: The Experiment; 2016.

22. Ehrensaft D. Boys will be girls, girls will be boys. Psychoanalytic Psychol. 2011;28(4):528–548.

23. Ehrensaft D. From gender Identity disorder to gender identity creativity: true gender self child therapy. J Homosex. 2012;59(3):337–356. [PubMed]

24. Malpas J. Between pink and blue: a multi-dimensional family approach to gender non-conforming children and their families. Fam Process. 2011;50(4):453–470. [PubMed]

25. Roberts L, Rosario M, Corliss HL, et al. Childhood gender nonconformity: a risk indicator for childhood abuse and post-traumatic stress in youth. Pediatrics. 2012;129:410–417. [PMC free article] [PubMed]

26. Wallace R, Russell H. Attachment and shame in gender-nonconforming children and their families: toward a theoretical framework for evaluating clinical interventions. Int J Transgenderism. 2013;14(3):113–126.

27. Olson KR, Key AC, Eaton NR. Gender cognition in transgender children. Psychol Sci. 2015;26(4):467–474. [PubMed]

28. Durwood L, McLaughlin KA, Olson KR. Mental health and self-worth in socially transitioned transgender youth. J Am Acad Child Adolesc Psychiatry. 2017;56(2):116–123. [PMC free article] [PubMed]

29. Rosenthal SM. Approach to the patient: transgender youth: endocrine considerations. J Clin Endocrinol Metab. 2014;99(12):4379–4389. [PubMed]

30. Vance SR, Ehrensaft D, Rosenthal SM. Psychological and medical care of gender nonconforming youth. Pediatrics. 2014;134(6):1184–1192. [PubMed]

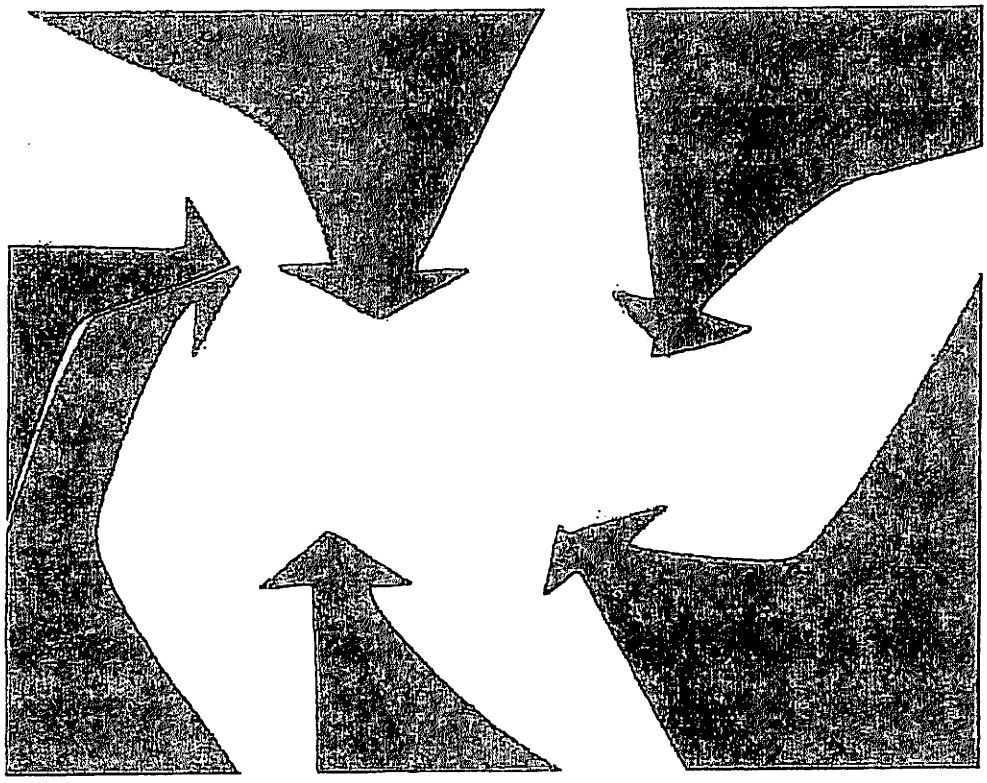
31. Ehrensaft D, Hastings J, Hsaio K. Ferility issues for youth: puberty blockers, hormones. Decision-making through the lens of family building options; World Professional Association for Transgender Health Symposia; Amsterdam. June 20, 2016.

32. Harris A. Gender as Soft Assembly. New York, NY: The Analytic Press; 2005.

33. deVries ALC, McGuire JK, Steensma TD, Wagenaar EC, Doreleijers TA, Cohen-Kettenis PT. Young adult psychological outcome after puberty suppression and gender reassignment. Pediatrics. 2014;134(4):696–704. [PubMed]

DAN KARASIC AND DIANE EHRENSAFT OPINION 07.06.15 07:00 AM

WE MUST PUT AN END TO GENDER CONVERSION THERAPY FOR KIDS



GETTY IMAGES

LEELAH ALCORN, A trans youth who endured conversion therapy, wrote in her suicide note, "The only way I will rest in peace is if one day transgender people aren't treated the way I was, they're treated like humans, with valid feelings and human rights."

Recently, a series of +++inset-left

WIRED OPINION
ABOUT

(https://www.wired.com/2015/06/big-problem-outlawing-gender-conversion-therapies/) pieces have been published in newspapers and magazines (including

~~therapy for LGBT youth, as well as the recently passed ban on conversion~~ therapy for LGBT youth in the province of Ontario, Canada home of some of the practice's defenders. Since the 1990s, major medical and mental health organizations have condemned conversion therapy for the purpose of changing a person's sexual orientation. The recent controversy is in response to efforts to also ban conversion therapy attempts to change gender identity and expression in children.

[opinion name="WIRED Opinion "]Dan Karasic is a Clinical Professor of Psychiatry at UCSF. He is on the Board of Directors of the World Professional Association for Transgender Health, and provides care for trans youth as a psychiatrist. Diane Ehrensaft is Director of Mental Health for the Child and Adolescent Gender Center at Benioff Children's Hospital and an Associate Professor of Pediatrics at UCSF. She is the author of *Gender Born, Gender Made: Raising Healthy Gender-Nonconforming Children.*

It is a misconception that this alternative to conversion therapy pushes children to transition to another gender.

Twitter

If we find ourselves steering someone in a direction that is *our* goal rather than theirs, we have failed in our job.

Twitter

#TRANSGENDER

VIEW COMMENTS

Expertise in Psychotherapy

An Elusive Goal?

Terence J. G. Tracey — *Arizona State University*
Bruce E. Wampold — *University of Wisconsin–Madison and Modum Bad Psychiatric Center, Vikersund, Norway*
James W. Lichtenberg — *University of Kansas*
Rodney K. Goodyear — *University of Houston and University of Redlands*

It has been argued that psychotherapy is a profession without any expertise (Shanteau, 1992). We examine the validity of this claim, reviewing the literature on expertise, clinical decision making, and psychotherapeutic outcome assessment, and find it a reasonable assessment. There is no demonstration of accuracy and skill that is associated with experience as a therapist. We posit that this absence of an expertise–experience relation is attributable to therapists' lack of access to quality outcome information regarding their interventions and an overreliance on fallible information-processing strategies even when such outcome information is available. The research on providing outcome feedback is reviewed, and although it does relate to client improvement, it has not been shown to be associated with any gains in therapist skill or expertise. We propose a model of outcome information usage and specific a priori hypothesis testing as a means of developing expertise.

Keywords: clinical decision making, clinical feedback, expertise

There is little debate regarding the efficacy and effectiveness of psychotherapy. Its benefits have been demonstrated repeatedly (Lambert & Ogles, 2004; Smith & Glass, 1977; Wampold, 2001a, 2001b). But as is the case for any human endeavor, the quality of psychotherapy varies across the people who provide it. Indeed, differences in outcomes among therapists have been detected in clinical trials (Baldwin & Imel, 2013; Crits-Christoph et al., 1991; Kim, Wampold, & Bolt, 2006), in naturalistic settings with therapists delivering a variety of treatments (Lutz, Leon, Martinovich, Lyons, & Stiles, 2007; Saxon & Barkham, 2012; Wampold & Brown, 2005), and in specialty clinics delivering a single evidence-based treatment (Laska, Smith, Wislocki, & Wampold, 2013). Clearly, some therapists are better than others—and, therefore, one could assert that there are some who are (or may be) "expert" therapists. But what *is* expertise in psychotherapy? How does it develop? What can be done to improve the expertise of therapists? We address these questions, but as will be demonstrated, expertise in psychotherapy is not a simple subject of inquiry.

In a review of expertise across professions, Shanteau (1992) identified several professions in which practitioners develop expertise, which he defined as increased quality of performance that is gained with additional experience. These professions, which demonstrate there is a relation between experience and professional skill, include astronomers, test pilots, chess masters, mathematicians, accountants, and insurance analysts. Shanteau also identified several professions for which expertise was *not* demonstrated, including psychiatrists, college admissions officers, court judges, personnel selectors, as well as clinical psychologists. He attributed the differences between the two types of professions to the predictability of their outcomes and the availability of quality feedback.

We argue that the tasks of psychotherapy make it difficult to obtain quality feedback about past actions, which in turn makes it difficult to develop expertise (Shanteau, 1992). As noted, psychotherapy is efficacious, and although psychotherapy is not unique with regard to difficulties in developing expertise, attention needs to be devoted to understanding the constraints on the development of therapist expertise, with the goal being that such attention will lead to better training of psychotherapists and improvement in the quality of mental health services.

In this article, we review the literature on expertise and how it applies to psychotherapy. We then focus on the constraints on skilled practice as well as the developing literature on feedback to the therapist about client progress. Finally, we discuss the conditions that are necessary for feedback to lead to expertise.

Our premise, like Shanteau's (1992) conclusion, is that over the course of one's professional practice as a psychotherapist, there is little development of expertise. We posit that this lack of expertise development (i.e.,

This article was published Online First January 6, 2014.
Terence J. G. Tracey, Department of Counseling and Counseling Psychology, Arizona State University; Bruce E. Wampold, Department of Counseling Psychology, University of Wisconsin—Madison, and Research Institute, Modum Bad Psychiatric Center, Vikersund, Norway; James W. Lichtenberg, Department of Psychology and Research in Education, University of Kansas; Rodney K. Goodyear, Department of Educational Psychology, University of Houston, and School of Education, University of Redlands.

We wish to acknowledge and thank the Otsego Group for its helpful and always engaging discussions on the subject matter of this article.

Correspondence concerning this article should be addressed to Terence J. G. Tracey, 446 Payne Hall, MC-0811, Arizona State University, Tempe, AZ 85287-0811. E-mail: terence.tracey@asu.edu

© 2014 American Psychological Association 0003-066X/14/$12.00
Vol. 69, No. 3, 218–229 DOI: 10.1037/a0035099

Terence J. G. Tracey

greater skill with greater experience) is attributable to the lack of information available to individual therapists regarding the outcomes of their interventions, the lack of adequate models about how psychotherapy produces benefits, and the difficulty of using the information that does exist to improve one's performance over time. We view the causes of this state of affairs as attributable to the current practice of psychotherapy and human information-processing difficulties and not as a failing that is unique or specific to psychotherapists as individuals.

Defining Expertise in Psychotherapy

One of the most obvious and enduring problems with respect to research on expertise has been the absence of a commonly accepted operational definition of expert performance—a problem that persists in considering expertise in psychotherapy. A therapist's expertise has been variously defined or understood in terms of his or her (a) reputation, (b) performance, or (c) client outcomes. Each of these conceptualizations is flawed. Reputation includes peer nomination, degree attainment, professional distinction such as diplomate status granted by the American Board of Professional Psychology, and overall amount of experience. Although these are desirable characteristics, their connection to improved performance and client outcomes is tenuous.

Performance has been defined via the demonstration of skill in performing psychotherapy tasks. For example, the APA Presidential Task Force on Evidence-Based Practice (2006; Sackett, Strauss, Richardson, Rosenberg, & Haynes, 2000) defined expertise as involving competence related to (a) assessment, diagnostic judgment, systematic case formulation, and treatment planning; (b) clinical decision making, treatment implementation, and monitoring of client progress; (c) interpersonal skills; (d) evaluation and use of research evidence; (e) understanding the influence of individual, cultural, and contextual differences; (f) understanding the influence of individual differences; and (g) having a cogent rationale for clinical strategies. These are desirable skills for a therapist to have, but they are difficult to define and assess, much less aggregate into an indicator of expertise. Further, an issue that arises is the distinction between expertise and competence. Although these two concepts are often used interchangeably, we think such usage fails to recognize that competence refers to capable performance, while expertise refers to expert performance that exceeds competence.

Performance has also been defined by treatment adherence, and the literature on the relation between treatment adherence and outcome is not clear. Webb, DeRubeis, and Barber (2010) found that, in general, competence in and adherence to clinical protocols (also known as "treatment manuals") appears to be unrelated to outcome overall, although these researchers did find modest support for relations with outcome when focusing on depression treatments. Also, research suggests that strict adherence to protocols might even attenuate therapeutic outcomes (Castonguay, Goldfried, Wiser, Raue, & Hayes, 1996; Henry, Strupp, Butler, Schacht, & Binder, 1993).

The final common definition of expertise is one based on client outcomes. Some have argued that the ultimate criterion of expertise is client outcome or client improvement (Wampold & Brown, 2005), with those therapists who produce the most improvement or best outcomes across clients being the experts. Although client outcome may be a reasonable criterion for evaluating expertise, especially considering accountability for services, it too is not without problems. Therapists do have an effect on outcomes of psychotherapy, although outcomes are due in large part to client variables, including severity of dysfunction, diagnosis, motivation (e.g., stage of change), social support, and resources (e.g., Bohart & Tallman, 2010; Groth-Marnat, Roberts, & Beutler, 2001). Some therapists work with more pathological or unmotivated clients than others and thus outcome scores will reflect this lack of comparability of client cases. In sum, there are many difficulties in the determination of individual expertise. However, here we focus on the expertise of the profession of psychotherapy and not on the determination of who is or is not an expert.

Given these limitations concerning the use simply of reputation, performance, or client outcome, we adopt the definition of expertise used by Shanteau (1992), which focuses on improvement over time. Expertise is improved performance that results from greater experience. Individuals should be able to use their practice to improve, and such improvement should manifest in better performance and outcomes.

As reasonable as this definition might be, the literature on experience fails to demonstrate that more experienced therapists are more effective than less experienced therapists (e.g., Hattie, Sharpley, & Rogers, 1984; Stein & Lambert, 1984, 1995; Wampold & Brown, 2005). Indeed,

Bruce E. Wampold

naturalistic studies have found that trainees attain client outcomes similar to those of licensed professionals (Beutler, 1997; Beutler et al., 2004; Budge et al., 2013; Laska et al., 2013; Minami et al., 2008; Okishi, Lambert, Nielsen, & Ogles, 2003), as do untrained college professors (Strupp & Hadley, 1979). So there does not appear to be a relation between professional experience and increased skill even when client outcome is used as the basis of skill definition. This lack of relation between experience and skill is the key basis for Shanteau's (1992) conclusion.

Barriers to Achieving Expertise in Psychotherapy

Several factors serve as barriers to achieving expertise in psychotherapy, including the cognitive and information processes of therapists, therapists' failure to engage routinely in deliberate practice, the inaccuracy of therapists' self-appraisals of their competence, and the lack of accurate feedback that affects learning. We address each of these barriers below.

Cognitive and Information-Processing Factors Affecting Expertise

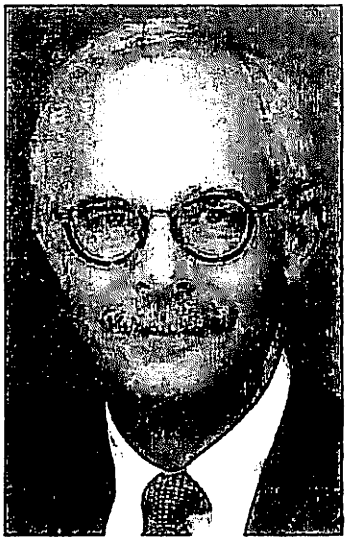
Cognitive differences between novices and experts (defined as those individuals who have better performance and outcomes) have been shown across a broad range of activities (domains), including computer programming, chess playing, teaching, driving a taxi cab, composing music, solving physics problems, deriving medical diagnoses, playing bridge, solving algebra word problems, solving economic problems, and judicial decision making (Chi, 2006; Feltovich, Prietula, & Ericsson, 2006). Compared with novices, experts have (a) the ability to perceive large meaningful patterns in their domain, (b) greater information-processing speed and accuracy, (c) superior long- and short-term memory, (d) the ability to see and represent a problem in their domain at a deeper (more meaningful) level than novices, (e) greater time spent understanding or analyzing problems qualitatively, and (f) stronger self-monitoring skills (i.e., they are better than novices at evaluating their own performance). Said another way, experts differ from novices by having (a) a larger knowledge base, (b) a better organized or structured knowledge base, and (c) greater proceduralization (automatic processing) of decision making.

For most situations, experts are advantaged by their larger and more structured knowledge base and the greater proceduralization of their decision making. But under some circumstances, expert performance is likely to be hampered (Lewandowsky & Thomas, 2009) and therefore poorer than that of nonexperts (Ericsson & Lehmann, 1996). This occurs when (a) basic-level information or nonintegrated information has to be retrieved, (b) individuals are forced to restructure their existing knowledge to incorporate new, incompatible information, or (c) existing knowledge has to be deliberately or consciously selected or new knowledge has to be created.

Although experienced therapists have more complete conceptualizations of clients than do novices, the accuracy of these conceptualizations has not been supported (Faust, 1991; Garb, 1998, 2005). In this regard, Dawes (1994) conceded that professional clinicians may make somewhat slightly better judgments in some circumstances than nonprofessionals, but these differences can generally be explained in terms of differences in such characteristics as intelligence and by the fact that people who have learned how to use valid diagnostic techniques employ them better than people who have not learned to use them. Further, once the rudiments of the techniques have been mastered, the accuracy of therapists' judgments generally does *not* increase with additional experience using them. Apparently, the selective advantage that professionals have over nonprofessionals lies in their mastering of the basics of valid techniques—the accuracy of those judgments being constrained by the accuracy of the techniques they employ. That is, even in the hands of experts, questionable techniques yield questionable predictions and judgments. In summary, clinical experience per se appears to do little to enhance accuracy of therapists' clinical judgments.

Failure to Engage Routinely in Deliberate Practice

A key aspect of professional expertise is that it is acquired through practice. More specifically, it is acquired through *deliberate* practice (Ericsson, 2006), which differs from mere exposure and repetition in several important ways. First, deliberate practice involves a well-defined, specific task that the learner seeks to master; second, task performance is followed by immediate feedback; third, there is opportunity for repetition; and fourth, learners must actively exploit the opportunity for improvement afforded by errors (Lewandowsky & Thomas, 2009, p. 143).



James W. Lichtenberg

But post-licensure therapists typically do not engage in this type of deliberate practice and, as we will discuss later, do not routinely obtain suitable feedback. As a consequence, Dawes (1994) was able to conclude, "The empirical data suggest that mental health professionals' accuracy of judgment does not increase with increasing experience" (p. 106).

Dawes's (1994) conclusion about the relationship between experience and expertise has not gone unchallenged. Perhaps the most recent of those challenges was Spengler et al.'s (2009) meta-analysis of 75 clinical judgment studies published between 1970 and 1996 that combined the experience of 4,607 clinicians. Spengler et al. found that the accuracy of clinical judgments *was* enhanced as a result of experience, although not by much (effect size: $d = 0.12$). However, therapists specifically trained in or with extensive clinical experience in a particular domain were no more or less accurate in that specific domain than were those without such specific training or experience. Further, the study did not tease apart the contribution of training from that of experience. Although Spengler et al.'s effect size was significant (and as they interpreted it, "not trivial"), Lichtenberg (2009) noted in his commentary on their article, "There is much history (of non-significant experience effects) for such a small effect size to overcome" (p. 413). Moreover, Huppert et al. (2001) found that there was a small association of overall therapy experience and outcomes with cognitive behavioral therapy with panic disorder patients. However, Huppert et al. also found among their 14 therapists that there was no relation of specific experience in cognitive behavioral therapy with outcome. So although there may be some effects of experience with outcome, these effects are small and not associated with the specific interventions used. What does change with experience is clinicians' confidence, which we discuss in the next section.

Inaccuracy of Therapists' Self-Appraisals of Competence

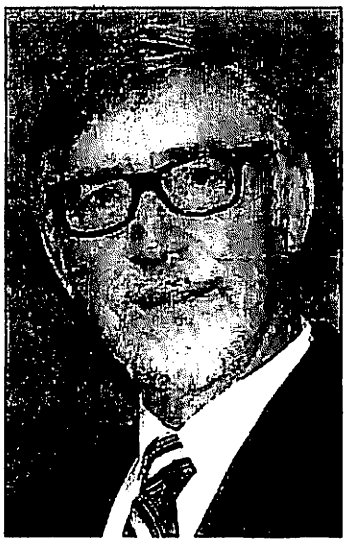
Clinicians have very unrealistic appraisals of their own competence. Walfish, McAlister, O'Donnell, and Lambert (2012) found, for example, that 25% of clinicians view themselves in the top 10% and that none viewed themselves as below average. Such perceptions of one's own competence are not unusual. There is a general tendency to fail to recognize one's own incompetence (Dunning, Johnson, Ehrlinger, & Kruger, 2003). Moreover, as Dawes (1994) pointed out, self-estimates of ability continue to grow with experience, even though actual ability does not. Such unwarranted growth in clinicians' confidence with experience has received substantial empirical documentation (Friedlander & Phillips, 1984; Goldberg, 1959, 1968; Oskamp, 1962, 1965; Rock, Bransford, Maist, & Morey, 1987).

People's confidence in their perception of others increases with experience and the richness of their mental representations, but this is not related to any increases in accuracy (Gill, Swann, & Silvera, 1998). The confidence that experienced therapists have in their predictions and the accuracy of these predictions are poorly related (Ægisdóttir et al., 2006; Garb, 1989, 2005; Goldberg, 1968; Spengler et al., 2009; Strasser & Gruber, 2004; Witteman & Van den Bercken, 2007; Witteman, Weiss, & Metzmacher, 2012). But to the extent that therapists *believe* they are growing more competent with experience, they are less likely to be motivated to take actions (e.g., obtain and use critical feedback) that would enhance their actual expertise (Pintrich, 2003).

Lack of Accurate Feedback

In a recent article on the conditions for intuitive expertise (i.e., the development of expertise that arises from continued practice and experience), Kahneman and Klein (2009) concluded, like Shanteau (1992), that expertise develops when two conditions exist: (a) The environment is predictable and with explicit outcomes, and (b) there is an opportunity to learn, based on quality information on the accuracy of past decisions and predictions (see also Ericsson, 2006). But the typical practice of psychotherapy meets neither of these conditions. As a result, psychotherapy is a context that provides little feedback regarding the accuracy of past clinical decisions and behaviors as well as client outcomes in general. In this context, we employ Hattie and Timperley's (2007) definition of feedback, which concerns information that reduces the discrepancy between current understandings or behaviors and those that are desired. We attribute this lack of quality information as a key reason for the difficulty in the development of expertise in psychotherapists. But this does not mean that there is a lack of information that could be utilized; there is, although too often it is flawed.

Much of the information therapists receive comes from clients: both what the clients report and what the



Rodney K. Goodyear

therapists are able to observe about their clients' functioning. Unfortunately, both types of information tend to be unreliable.

The well-known "Barnum effect" (Meehl, 1956) can be used to exemplify problems with client reports. Clients have been found to be willing to accept almost any interpretation as accurate (C. R. Snyder, Shenkel, & Lowrey, 1977), so a client's report of therapist accuracy may be misleadingly affirmative. The literature on the predictive validity of interviewing has long demonstrated the weakness of interview-based decisions and predictions (Carroll, Wiener, Coates, Galegher, & Alibrio, 1982; DeVaul et al., 1987; Milstein, Wilkinson, Burrow, & Kessen, 1981)—a finding that has been shown to generalize to clinical practice (Garb, 1998; Oskamp, 1965). Using standardized interview formats or valid psychological assessment has been shown to provide better-quality information and enhanced validity of predictions.

Another problem related to information quality occurs with respect to the manner in which therapists assess treatment outcomes. Certainly therapists routinely make judgments about how their clients are functioning at termination and about the likelihood of their continuing to do well outside of therapy. From these judgments, they make outcome attributions and derive representations of why the changes occurred in therapy. This type of model creation is central to the work of therapists. In fact, Voss and his colleagues (Voss, Greene, Post, & Penner, 1983; Voss & Post, 1988; Voss, Tyler, & Yengo, 1983) have found that experts in any domain are better able than novices to build coherent and persuasive explanations for their solutions to problems.

But therapists rarely, if ever, test their models. It is uncommon enough for clinicians to get information on how their clients are doing during their therapy, but it is rarer still for them to do so after termination. All that therapists generally have is the report of clients in the last few sessions. It is generally understood and accepted that these tend to be overly positive and unrepresentative outcome evaluations—a phenomenon that is sufficiently common that it has been named the Hello-Goodbye effect (Hathaway, 1948), in which client assessments made at termination tend to be inflated relative to those obtained shortly thereafter. Unfortunately, as far as feedback goes, the only definitive outcome information therapists may get is that which occurs when a client later returns for more treatment.

The problem, however, goes beyond the simple availability of quality information and gets to therapists' intentionality in seeking and using what information is available to them. It is instructive to consider conclusions Miller, Hubble, and Duncan (2008) have drawn from studying the very best therapists, who are defined as those with the highest outcomes. They found that these therapists "without exception possess a keen 'situational awareness': they are observant, alert and attentive. They compare new information constantly with what they already know" (Miller et al., 2008, p. 19). So we see the failure of therapists getting better with experience as related to cognitive processing issues and lack of quality outcome information. But these issues can be overcome.

Methods to Increase Expertise

Studies of different interventions and large-scale insurance data systems provide abundant aggregate client outcome information. But there is relatively little reporting of outcome information to individual therapists and even less reporting of one clinician's client outcomes relative to that of other clinicians. We see the issue of obtaining information on how the client is progressing as crucial in the development of clinical skill. Further, the ways in which therapists use this information for hypothesis development and testing is essential. In what follows we report promising practices even as we consider factors that temper their overall effectiveness.

Systematic Feedback on Client Progress

One source of feedback is the provision of client progress information to the therapist, an idea attributed to Ken Howard and colleagues (Howard, Moras, Brill, Martinovich, & Lutz, 1996) and developed and tested by both Michael Lambert and colleagues (Lambert, Hansen, & Finch, 2001) and Scott Miller and colleagues (Miller, Duncan, & Hubble, 2005). The systems for providing feedback about client progress to therapists (e.g., Barkham, Hardy, & Mellor-Clark, 2010; Duncan, Miller, Wampold, & Hubble, 2010; Lambert, 2010; Lambert et al., 2001; Lambert, Harmon, Slade, Whipple, & Hawkins, 2005; Miller, Duncan, & Hubble, 2005) involve several components.

- First, a measure is used to assess the functioning of the client periodically during therapy. Typically that measure is global, as opposed to disorder-specific, in that it assesses well-being, role functioning, interpersonal func-

tioning, as well as symptom distress, issues that are applicable across disorders.

- Second, normative trajectories of client progress are derived from the progress of large samples of clients.
- Finally, individual client progress is compared to the normative trajectories, and feedback is provided to the therapist about client progress relative to the norms, usually adjusted for "case mix." For example, the therapist may receive a "red dot" or a "red light" if client progress is in the lowest percentiles—for example, if the client progress is less than the progress of 25% of clients with a comparable number of sessions and similar initial severity (Lambert et al., 2005; Miller, Duncan, & Hubble, 2005). Clients in this category have been labeled "not on track" (NOT) cases or signal cases. Because therapists typically do not recognize deteriorating cases (Hannan et al., 2005), providing therapists feedback on NOT cases demonstrates the discrepancy between the therapist's view of client progress and actual client progress, which is hypothesized to be an important aspect of feedback interventions (Sapyta, Riemer, & Bickman, 2005). Therapists also receive feedback about clients who are progressing normally (i.e., their client's change is at or above average) or are at risk for a poor outcome (i.e., below average but not in the bottom quartile).

There is sufficient evidence to conclude that providing this feedback to therapists positively affects outcome (Lambert & Shimokawa, 2011; Shimokawa, Lambert, & Smart, 2010). In randomized trials, clients in conditions in which their therapist received feedback had better outcomes than clients whose therapists did not receive feedback, although the effects were achieved primarily by reducing the proportion of clients who deteriorated—that is, by reducing the rate of failures in NOT cases (Lambert & Shimokawa, 2011; Shimokawa et al., 2010).

This is good news, but from our perspective and with particular regard to the development of therapist expertise, it is important to know what therapists who receive feedback do to improve their clients' outcomes and to reduce the rate of deterioration. It may well be that therapists receiving feedback learn skills that are generalizable to other clients and thus become more competent therapists (i.e., develop expertise), or it may be that therapists simply pay more attention to NOT cases than they were doing previously, may query the client about progress (but only when prompted by a red dot or red signal), or may encourage the client to respond more positively. Unfortunately, there has been no research aimed at identifying what therapists do in response to feedback or what the feedback affects.

In this regard, studies examining feedback by and large have examined the outcomes of *clients* rather than the behavior of *therapists*. Consequently, it is difficult to know what effect feedback has on therapists and specifically on therapist expertise. Indeed, the usual practice is to provide feedback about clients rather than feedback about therapists. This difference is important. In the latter situation, therapists would receive feedback about their performance *relative to the performance of other therapists*, which would then give therapists opportunity to understand the discrepancy between their belief about their competence and their actual competence. Although some systems are set up to provide therapists with feedback about their relative performance across clients, there has been no research focused on feedback to therapists about their performance relative to other therapists.

Utility of systematically obtained client feedback. Feedback to therapists would be important for two reasons. First, outcome differences across therapists seem to be robust in practice settings (Laska et al., 2013; Lutz et al., 2007; Saxon & Barkham, 2012; Wampold & Brown, 2005; see Baldwin & Imel, 2013, for a review). Some therapists consistently attain better outcomes than others, and therefore feedback should help the less effective therapists improve their performance. It would be informative to know whether feedback about the progress of individual clients or feedback about therapists' relative performance is more helpful to therapists of different effectiveness levels. For example, it could be that feedback about clients would be more helpful to more effective therapists because these therapists may be more open to client information and skilled in using that information. Conversely, more effective therapists may already be aware of client progress through their interaction with clients, and it may be the less effective therapists, who need information about client progress, who benefit from feedback. Because therapists generally believe that they are effective (i.e., above average; Walfish et al, 2012), feedback concerning therapists' relative efficacy would make apparent the discrepancy between self-assessment and actual performance, which represents a fruitful focus for future research.

Necessary but insufficient for developing expertise. We contend that feedback about client progress, either with regard to individual clients or with regard to therapists' relative effectiveness, is necessary but not sufficient to develop expertise. Feedback about client progress provides no information about what actions are necessary to improve performance. To develop expertise, feedback needs to be specific to the important components of psychotherapy. That observation, unfortunately, reveals exactly why developing expertise in psychotherapy is so elusive. Generally speaking, there is little agreement about models of psychotherapy that would form the basis of focusing on component processes.

It should be clear that there are very different treatments, with very different protocols, for any given disorder, some of which have been identified as empirically supported or research based (http://www.apa.org/divisions/div12/cppi.html). According to this perspective, expertise is defined for a specific treatment as suggested by Waltz, Addis, Koerner, and Jacobson (1993). However, because all treatments that are intended to be therapeutic seem to be approximately equally effective (Laska, Gurman, & Wampold, in press; Wampold, 2001b; Wampold et al., 1997) it may well be that adherence to a protocol for a specific treatment may involve focusing on a component that does not produce better outcomes. Such speculation is

supported by the fact that protocol adherence generally does not seem to be related to outcome. In this regard, a meta-analysis by Webb et al. (2010) found no relation between treatment adherence and client outcome except for a small effect for depression treatments (and, interestingly, many very different treatments have been found to be efficacious for the treatment of depression). Indeed Boswell et al. (2013) have found that both adherence and competence deteriorate over time—the opposite of what one would expect if the expertise–experience relation were defined using adherence and competence as the criteria for expertise. At best, current practice supports the need for provision of normative information in every case about (a) client progress over treatment as well as (b) client outcome.

The presence of feedback information does not appear to lead to the development of expertise. A meta-analysis of the general use of feedback interventions across all domains, not just clinical domains, reveals a small positive effect size, but the results are quite variable, with roughly one third of the interventions producing negative effects (Kluger & DeNisi, 1996).

Planful Application of Feedback Information

There are several recommendations in the literature for improving our clinical decision making, which serves as one the cornerstones of clinical intervention (e.g., Arkes, 1981; Dumont, 1991; Faust, 1991; Garb, 1998; Garb & Boyle, 2003; Goldberg, 1991; Salovey & Turk, 1991; Tracey & Rounds, 1999; Wierzbicki, 1993), but these recommendations apparently have had little effect (Lilienfeld, Lynn, & Lohr, 2003; Lilienfeld, Ritschel, Lynn, Cautin, & Latzman, 2013). As demonstrated by Lewandowsky, Ecker, Seifert, Schwarz, and Cook (2012), it is difficult to dispel mistaken or inaccurate information and conclusions, and so this state of affairs is to be somewhat expected.

Recommended procedures for enhancing clinical decision making and practice include (a) adopting a Bayesian approach by looking at base rates and the predictability of behavior, (b) obtaining quality information (e.g., relying on valid measures rather than impressions), (c) relying less on memory, (d) recognizing personal biases and their effects, (e) being aware of regression to the mean where less extreme behavior follows extreme behavior, and (f) adopting a disconfirming, scientific approach to practice. We see merit in each of these recommendations but wish to focus specifically on the last of these, that of adopting a disconfirming, scientific method. We see this as the most central recommendation relative to obtaining expertise. Three issues are of particular relevance to adopting a disconfirming, scientific approach: (a) overuse of confirmatory bias, (b) overuse of hindsight bias, and (c) failure to engage in specific hypothesis testing. Seeking to alter each of these issues is crucial in gaining expertise.

Adopt a disconfirmatory approach. People tend to seek confirmatory information concerning their beliefs (Davies, 2003; Granberg & Brent, 1983; Sears & Whitley, 1973) by seeking out and attending to information that confirms their concepts (Aldashev, Carletti, & Righi, 2011; Greenwald, Pratkanis, Leippe, & Baumgardner, 1986; Nisbett & Ross, 1980; M. Snyder & Campbell, 1980). The effect is that only partial evidence is perceived. This tendency to seek confirmation also has been demonstrated in therapists (Haverkamp, 1993; Strohmer, Shivy, & Chiodo, 1990). If therapists believe something to be true, the natural approach is to look for evidence that confirms, rather than tests, this belief.

Such a confirmatory approach leads to biased information searches and a high probability of incorrect conclusions. A wiser approach would be to adopt a disconfirming approach. Using this approach, the individual specifies what information would be needed to render the belief wrong and then seeks such disconfirming information. Such disconfirming information searches yield more and better-quality information and thus provide a more accurate base for decision making (Aldashev et al., 2011; Davies, 2003).

Avoid hindsight bias. A second issue pertaining to the need to adopt a disconfirming, scientific approach is the avoidance of hindsight bias (Wedding & Faust, 1989) in clinical practice. Hindsight bias is akin to "Monday morning quarterbacking," in which everyone knows the optimal play after the fact. Such post hoc construction of models and explanations creates an illusion of learning in which individuals believe these counterfactual models to be accurate and that they knew about them all along (Roese & Vohs, 2012).

Although post hoc models do have some utility in furthering our understanding of clients, this learning is illusory in that it is never really examined. There is no a priori testing of hypotheses that emanate from therapists' understanding of the client. Hindsight bias is used frequently in practice (Arkes, Faust, Guilmette, & Hart, 1988; Arkes, Wortmann, Saville, & Harkness, 1981; Fischhoff, 1975) and results, unfortunately, in a lack of information gathering. If new information arises, then it is easily incorporated into clinical conceptualizations after the fact (Roese & Vohs, 2012). As a result, clinicians are rarely wrong because they have never really tested the validity of their beliefs. Hindsight bias is especially salient with respect to getting and dealing with feedback on client outcomes. If a therapist learns that a client had lower outcomes than thought based on termination comments, it becomes relatively easy to construct an explanation. But the key issue is that this explanation then needs to be explicitly tested on new clients. If the therapist made strong predictions about the outcome and then found out that the actual outcome differed, he or she would then have clear information on the need to change the clinical formulation. The probability of subsequent hypothesis specification would probably be greater in this case, and the likelihood of testing these hypotheses with future clients should also increase. A similar process could be used in ongoing work with continuing clients. Generating explicit predictions about client progress and then receiving feedback on the extent of that progress serves as an explicit guard against hindsight bias.

Explicitly test hypotheses. The final aspect of our advocating the adoption of a disconfirming, scientific approach is the clear specification and evaluation of clinical

hypotheses. This evaluation requires collecting disconfirming information relative to future, not past, behavior. We argue that learning occurs by specifying and then testing specific, a priori, empirically verifiable hypotheses. If a therapist claims that a client will perform certain behaviors based upon his or her conceptualization of the client, the therapist needs to propose specific hypotheses, focusing on specific evidence that would disconfirm the hypothesis (i.e., alternative hypotheses), and then set up conditions to test the hypothesis. Doing this in a prospective manner enables the acquisition of information on the validity of the models and one's hypotheses. Evidence of inaccuracy is unequivocal and necessitates a change in the understanding of the client.

In this regard, we are suggesting that psychotherapists fail to develop clinical expertise because of their failure to adopt a disconfirming scientific process in practice even when there is quality feedback information such as information on outcomes. As an oversimplified example, a clinician with poorer than expected outcomes might hypothesize that the relative ratio of focus on positive to negative content could be a key variable accounting for these results. The clinician could then systematically vary the content of the session to see its effect on progress and outcomes. What also would be needed is a specification of what should occur in session should this hypothesis be either (a) correct or (b) incorrect. Besides the value of the specification of a hypothesis, it is the addition of the disconfirmation that makes this strategy valuable.

A key aspect of the disconfirming, scientific method is the generation of testable hypotheses. Although it is not difficult to proffer hypotheses—indeed, it is done frequently—a key requirement is that the hypotheses be embedded in a clearly articulated model of client processing and behavior. The tests of the hypotheses generated by the model thus provide information on the validity of the model. It is this generation of specific hypotheses, confirmed by experience in deliberate practice, that, we believe, forms the basis of the development of psychotherapeutic expertise. A recent study found that therapist perceptions of professional self-doubt were positively related to therapy outcome (Nissen-Lie, Monsen, Ulleberg, & Rønnestad, 2012). Although this professional self-doubt may not explicitly comprise our proposed scientific testing, it does appear to encompass a critical evaluation of one's work from a disconfirming stance. Williams, Dunning, and Kruger (2013) have demonstrated that inflated self-assessments of performance are associated with rational, rule-based methods relative to more variable approaches. Given the high occurrence of confirmatory approaches, it is likely that this rule-based rational approach includes a good deal of confirmatory bias. So being more pessimistic regarding one's client's outcome may be an asset in that it may be associated with the application of more alternative explanations than the rule-based confirmatory method.

Deliberate practice. Miller (Miller, Duncan, Sorrell, & Brown, 2005; Miller, Hubble, Chow, & Seidel, 2013) has recommended deliberate practice as the means by which clinical expertise can be attained. Deliberate practice is defined as the explicit setting aside of private time to review one's behavior and outcome feedback, developing plans for improvement, and then following through on these. The expertise literature has demonstrated that such deliberate practice is associated with the attainment of expertise in a variety of domains (Ericsson, 2009). However, recent research has demonstrated, at least with chess masters, that deliberate practice is necessary, but not sufficient, for the development of expertise (Campitelli & Gobet, 2011). The specific type of deliberate practice matters. Fischer, Fischer, Weisweiler, and Frey (2010) found that confirmatory bias was greatest in conditions of deliberate cognitive analysis and intuitive and gut feelings. Conditions of distraction (i.e., doing other tasks) resulted in the least confirmatory bias. So deliberate rational processing alone (or intuitive processing alone) does not result in disconfirmatory processing. Given the literature on disconfirmatory or alternative hypothesis testing, it would be expected that this reflective method of deliberate processing coupled with alternate hypothesis generation would also result in less confirmatory bias. We agree with Miller, Duncan, et al. (2005; Miller et al., 2013) that deliberate practice is essential, but we add that this practice should be of a particular form, that of setting aside explicit time to generate a priori alternative or disconfirmatory hypotheses and then testing them explicitly. Simply reflecting in a deliberate manner on feedback information is insufficient.

Conclusion

Shanteau (1992) claimed that the practice of psychotherapy does not have an expertise base in that there is little relation between experience and gains in professional skill. We have discussed several aspects of psychotherapy that make the development of expertise as a therapist particularly difficult. Essentially, psychotherapy is a process about which the therapist receives little explicit and valid feedback about what actions are productive of a therapeutic outcome.

Notwithstanding the above difficulty, there is extensive evidence that psychotherapy is effective. As well, there are documented differences among the outcomes achieved by therapists—some therapists consistently achieve better outcomes than others. Thus, although it appears that there exists such a thing as expertise, little is known about what differentiates the more effective therapists from others; certainly it does not appear to be the type of therapy delivered or the experience of the therapist (Beutler et al., 2004). What has emerged is that more effective therapists appear to be able to form working alliances across a range of clients (Baldwin, Wampold, & Imel, 2007) and have a greater level of facilitative skills (Anderson, Ogles, Patterson, Lambert, & Vermeersch, 2009). But this information provides little that is actionable to facilitate the development of expertise. Clearly, more research about the process and outcome of psychotherapy is needed (see Kazdin, 2008), including what characterizes expert therapists with better outcomes, because it is clear

that better outcomes do not emerge as a function of experience.

Despite the barriers to developing expertise in psychotherapy, there is much clinicians can do. Increasingly, there are reliable benchmarks for various disorders (e.g., see Minami et al., 2008, with regard to depression) to which therapists can compare the progress made by their clients. Therapists can use feedback about client progress to adjust therapy to achieve optimal outcomes and to compare their outcomes to those of other therapists. Therapists, particularly those who are underperforming relative to other therapists, can seek to improve.

In this regard, therapists may need to augment their general therapeutic competence or they may need to be trained to provide particular evidence-based psychological treatments, depending on the reasons for their relatively poor performance. It is crucial that therapists obtain quality information about both client and therapist outcomes if they are to establish expertise. However, as we have argued, outcome information alone, even if of high quality, does not ensure that expertise will develop. Cognitive heuristics, especially hindsight bias, can minimize the impact of outcome information on future practice. To benefit from quality information, therapists are encouraged to adopt a prospective testing of hypotheses, where the outcome information serves as the criterion. It is our hypothesis that adopting such a disconfirming, scientific approach to practice will result in expertise gains among therapists.

REFERENCES

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., & Nichols, C. S. (2006). The Meta-Analysis of Clinical Judgment Project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34,* 341–382. doi:10.1177/0011000005285875

Aldashev, G., Carletti, T., & Righi, S. (2011). Follies subdued: Informational efficiency under adaptive expectations and confirmatory bias. *Journal of Economic Behavior & Organization, 80,* 110–121. doi:10.1016/j.jebo.2011.03.001

Anderson, T., Ogles, B. M., Patterson, C. L., Lambert, M. J., & Vermeersch, D. A. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist success. *Journal of Clinical Psychology, 65,* 755–768. doi:10.1002/jclp.20583

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61,* 271–285. doi:10.1037/0003-066X.61.4.271

Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology, 49,* 323–330. doi:10.1037/0022-006X.49.3.323

Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology, 73,* 305–307. doi:10.1037/0021-9010.73.2.305

Arkes, H. R., Wortmann, R. L., Saville, P., & Harkness, A. R. (1981). The hindsight bias among physicians weighting the likelihood of diagnosis. *Journal of Applied Psychology, 66,* 252–254. doi:10.1037/0021-9010.66.2.252

Baldwin, S. A., & Imel, Z. (2013). Therapist effects. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavioral change* (6th ed., pp. 258–297). New York, NY: Wiley.

Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance–outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology, 75,* 842–852. doi:10.1037/0022-006X.75.6.842

Barkham, M., Hardy, G. E., & Mellor-Clark, J. (Eds.). (2010). *Developing and delivering practice-based evidence: A guide for the psychological therapies.* Chichester, England: Wiley Blackwell. doi:10.1002/9780470687994

Beutler, L. E. (1997). The psychotherapist as a neglected variable in psychotherapy: An illustration by reference to the role of therapist experience and training. *Clinical Psychology: Science and Practice, 4,* 44–52. doi:10.1111/j.1468-2850.1997.tb00098.x

Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., & Wong, E. (2004). Therapist variables. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 227–306). New York, NY: Wiley.

Bohart, A. C., & Tallman, K. (2010). Clients: The neglected common factor in psychotherapy. In B. L. Duncan, S. D. Miller, B. E. Wampold, & M. A. Hubble (Eds.), *The heart and soul of change: Delivering what works in therapy* (2nd ed., pp. 83–111). Washington, DC: American Psychological Association. doi:10.1037/12075-003

Boswell, J. F., Gallagher, M. W., Sauer-Zavala, S. E., Bullis, J., Gorman, J. M., Shear, M. K., . . . Barlow, D. H. (2013). Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for panic disorder. *Journal of Consulting and Clinical Psychology, 81,* 443–454. doi:10.1037/a0031437

Budge, S. L., Owen, J. J., Kopta, S. M., Minami, T., Hansen, M. R., & Hirsch, G. (2013). Differences among trainees in client outcomes associated with the phase model of change. *Psychotherapy, 50,* 150–157. doi:10.1037/a0029565

Campitelli, G., & Gobet, F. (2011). Deliberate practice: Necessary but not sufficient. *Current Directions in Psychological Science, 20,* 280–285. doi:10.1177/0963721411421922

Carroll, J. S., Wiener, R. L., Coates, D., Galegher, J., & Alibrio, J. J. (1982). Evaluation, diagnosis, and prediction in parole decision-making. *Law & Society Review, 17,* 199–228. doi:10.2307/3053536

Castonguay, L. G., Goldfried, M. R., Wiser, S., Raue, P. J., & Hayes, A. M. (1996). Predicting the effect of cognitive therapy for depression: A study of unique and common factors. *Journal of Consulting and Clinical Psychology, 64,* 497–504. doi:10.1037/0022-006X.64.3.497

Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, R. R. Hoffman, & P. J. Feltovich (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 21–30). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511816796.002

Crits-Christoph, P., Baranackie, K., Kurcias, J. S., Carroll, K., Luborsky, L., McLellan, T., & Zitrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 1,* 81–91. doi:10.1080/10503309112331335511

Davies, M. F. (2003). Confirmatory bias in the evaluation of personality descriptions: Positive test strategies and output interference. *Journal of Personality and Social Psychology, 85,* 736–744. doi:10.1037/0022-3514.85.4.736

Dawes, R. M. (1994). *House of cards: Psychology and psychiatry built on myth.* New York, NY: Free Press.

DeVaul, R. A., Jervey, F., Chappell, J. A., Carver, P., Short, B., & O'Keefe, S. (1987). Medical school performance of initially rejected students. *JAMA: Journal of the American Medical Association, 257,* 47–51. doi:10.1001/jama.1987.03390010051027

Dumont, F. (1991). Expertise in psychotherapy: Inherent liabilities of becoming experienced. *Psychotherapy: Theory, Research, Practice, Training, 28,* 422–428. doi:10.1037/0033-3204.28.3.422

Duncan, B. L., Miller, S. D., Wampold, B. E., & Hubble, M. A. (Eds.). (2010). *The heart and soul of change: Delivering what works in therapy* (2nd ed.). Washington, DC: American Psychological Association. doi:10.1037/12075-000

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12,* 83–87. doi:10.1111/1467-8721.01235

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–703). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511816796.038

Ericsson, K. A. (2009). Enhancing the development of professional per-

formance: Implications from the study of deliberate practice. In K. A. Ericsson (Ed.), *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments* (pp. 405–431). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511609817.022

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47,* 273–305. doi:10.1146/annurev.psych.47.1.273

Faust, D. (1991). What if we had really listened? Present reflections on altered pasts. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology: Vol. 1. Matters of public interest* (pp. 185–216). Minneapolis, MN: University of Minnesota Press.

Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, R. R. Hoffman, & P. J. Feltovich (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 41–67). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511816796.004

Fischer, P., Fischer, J., Weisweiler, S., & Frey, D. (2010). Selective exposure to information: How different modes of decision making affect subsequent confirmatory information processing. *British Journal of Social Psychology, 49,* 871–881. doi:10.1348/014466610X499668

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 288–299. doi:10.1037/0096-1523.1.3.288

Friedlander, M. L., & Phillips, S. D. (1984). Preventing anchoring errors in clinical judgment. *Journal of Consulting and Clinical Psychology, 52,* 366–371. doi:10.1037/0022-006X.52.3.366

Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105,* 387–396. doi:10.1037/0033-2909.105.3.387

Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment.* Washington, DC: American Psychological Association. doi:10.1037/10299-000

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1,* 67–89. doi:10.1146/annurev.clinpsy.1.102803.143810

Garb, H. N., & Boyle, P. A. (2003). Understanding why some clinicians use pseudoscientific methods: Findings from research on clinical judgment. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 17–38). New York, NY: Guilford Press.

Gill, M. J., Swann, W. B., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology, 75,* 1101–1114. doi:10.1037/0022-3514.75.5.1101

Goldberg, L. R. (1959). The effectiveness of clinician's judgments: The diagnosis of organic brain damage from the Bender-Gestalt. *Journal of Consulting Psychology, 23,* 25–33. doi:10.1037/h0048736

Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgment. *American Psychologist, 23,* 483–496. doi:10.1037/h0026206

Goldberg, L. R. (1991). Human mind versus regression equation: Five contrasts. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology: Vol. 1. Matters of public interest* (pp. 173–184). Minneapolis, MN: University of Minnesota Press.

Granberg, D., & Brent, E. (1983). When prophecy bends: The preference-expectation link in U.S. presidential elections, 1952–1980. *Journal of Personality and Social Psychology, 45,* 477–491. doi:10.1037/0022-3514.45.3.477

Greenwald, A. G., Pratkanis, A. R., Leippe, M. R., & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review, 93,* 216–229. doi:10.1037/0033-295X.93.2.216

Groth-Marnat, G., Roberts, R., & Beutler, L. E. (2001). Client characteristics and psychotherapy: Perspectives, support, interactions, and implications for training. *Australian Psychologist, 36,* 115–121. doi:10.1080/00050060108259643

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology: In Session, 61,* 1–9.

Hathaway, S. R. (1948). Some considerations relative to nondirective counseling as therapy. *Journal of Clinical Psychology, 4,* 226–231. doi:10.1002/1097-4679(194807)4:3<226::AID-JCLP2270040303>3.0.CO;2-V

Hattie, J. A., Sharpley, C. F., & Rogers, H. F. (1984). Comparative effectiveness of professional and paraprofessional helpers. *Psychological Bulletin, 95,* 534–541. doi:10.1037/0033-2909.95.3.534

Hattie, J. A., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77,* 81–112. doi:10.3102/003465430298487

Haverkamp, B. E. (1993). Confirmatory bias in hypothesis testing for client-identified and counselor self-generated hypotheses. *Journal of Counseling Psychology, 40,* 303–315. doi:10.1037/0022-0167.40.3.303

Henry, W. P., Strupp, H. H., Butler, S. F., Schacht, T. E., & Binder, J. (1993). Effects of training in time-limited psychotherapy: Changes in therapist behavior. *Journal of Consulting and Clinical Psychology, 61,* 434–440. doi:10.1037/0022-006X.61.3.434

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist, 51,* 1059–1064. doi:10.1037/0003-066X.51.10.1059

Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2001). Therapists, therapists' variables, and cognitive behavioral therapy outcomes in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology, 69,* 747–755. doi:10.1037/0022-006X.69.5.747

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64,* 515–526. doi:10.1037/a0016755

Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63,* 146–159. doi:10.1037/0003-066X.63.3.146

Kim, D. M., Wampold, B. E., & Bolt, D. M. (2006). Therapist effects in psychotherapy: A random effects modeling of the NIMH TDCRP data. *Psychotherapy Research, 16,* 161–172. doi:10.1080/10503300500264911

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119,* 254–284. doi:10.1037/0033-2909.119.2.254

Lambert, M. J. (2010). *Prevention of treatment failure: The use of measuring, monitoring, and feedback in clinical practice.* Washington, DC: American Psychological Association. doi:10.1037/12141-000

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology, 69,* 159–172. doi:10.1037/0022-006X.69.2.159

Lambert, M. J., Harmon, C., Slade, K., Whipple, J. L., & Hawkins, E. J. (2005). Providing feedback to psychotherapists on their patients' progress: Clinical results and practice suggestions. *Journal of Clinical Psychology, 61*(2), 165–174. doi:10.1002/jclp.20113

Lambert, M. J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 139–193). New York, NY: Wiley.

Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 203–223). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199737208.003.0010

Laska, K. M., Gurman, A. S., & Wampold, B. E. (in press). Expanding the lens of evidence based practice in psychotherapy: A common factors perspective. *Psychotherapy.*

Laska, K. M., Smith, T. L., Wislocki, A. P., & Wampold, B. E. (2013). Uniformity of evidence-based treatments in practice? Therapist effects in the delivery of cognitive processing therapy for PTSD. *Journal of Counseling Psychology, 60,* 31–41. doi:10.1037/a0031294

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13,* 106–131. doi:10.1177/1529100612451018

Lewandowsky, S., & Thomas, J. L. (2009). Expertise: Acquisition, limitations, and control. *Reviews of Human Factors and Ergonomics, 5,* 140–165. doi:10.1518/155723409X448044

Lichtenberg, J. W. (2009). Comment: Effects of experience on judgment accuracy. *The Counseling Psychologist, 37,* 410–415. doi:10.1177/0011000008330829

Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (2003). Science and pseudoscience in clinical psychology: Initial thoughts, reflections, and considerations. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 1–14). New York, NY: Guilford Press.

Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2013). Why many clinical psychologists are resistant to evidence-based practice: Root causes and constructive remedies. *Clinical Psychology Review, 33,* 883–900. doi:10.1016/j.cpr.2012.09.008

Lutz, W., Leon, S. C., Martinovich, Z., Lyons, J. S., & Stiles, W. B. (2007). Therapist effects in outpatient psychotherapy: A three-level growth curve approach. *Journal of Counseling Psychology, 54,* 32–39. doi:10.1037/0022-0167.54.1.32

Meehl, P. E. (1956). Wanted—A good cookbook. *American Psychologist, 11,* 263–272. doi:10.1037/h0044164

Miller, S. D., Duncan, B. L., & Hubble, M. A. (2005). Outcome-informed clinical work. In J. C. Norcross & M. R. Goldfried (Eds.), *Handbook of psychotherapy integration* (2nd ed., pp. 84–102). New York, NY: Oxford University Press.

Miller, S. D., Duncan, B. L., Sorrell, R., & Brown, J. (2005). The partners for change outcome management system. *Journal of Clinical Psychology, 61,* 199–208. doi:10.1002/jclp.20111

Miller, S. D., Hubble, M. A., Chow, D. L., & Seidel, J. A. (2013). The outcome of psychotherapy: Yesterday, today, and tomorrow. *Psychotherapy, 50,* 88–97. doi:10.1037/a0031097

Miller, S. D., Hubble, M., & Duncan, B. (2008). Supershrinks: What is the secret of their success? *Psychotherapy in Australia, 14*(4), 14–22.

Milstein, R. M., Wilkinson, L., Burrow, G. N., & Kessen, W. (1981). Admission decisions and performance during medical school. *Journal of Medical Education, 56,* 77–82.

Minami, T., Wampold, B. E., Serlin, R. C., Hamilton, E., Brown, G. S., & Kircher, J. (2008). Benchmarking the effectiveness of psychotherapy treatment for adult depression in a managed care environment: A preliminary study. *Journal of Consulting and Clinical Psychology, 76,* 116–124. doi:10.1037/0022-006X.76.1.116

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* New York, NY: Prentice-Hall.

Nissen-Lie, H. A., Monsen, J. T., Ulleberg, P., & Rønnestad, M. H. (2012). Psychotherapists' self-reports of their interpersonal functioning and difficulties in practice as predictors of patient outcome. *Psychotherapy Research, 23,* 86–104. doi:10.1080/10503307.2012.735775

Okiishi, J., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An empirical analysis of therapist effects. *Clinical Psychology and Psychotherapy, 10,* 361–373. doi:10.1002/cpp.383

Oskamp, S. (1962). The relationship of clinical experience and training methods to several criteria of clinical prediction. *Psychological Monographs: General and Applied, 76*(28), 1–27. doi:10.1037/h0093849

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology, 29,* 261–265. doi:10.1037/h0022125

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95,* 667–686. doi:10.1037/0022-0663.95.4.667

Rock, D. L., Bransford, J. D., Maisto, S. A., & Morey, L. (1987). The study of clinical judgment: An ecological approach. *Clinical Psychology Review, 7,* 645–661. doi:10.1016/0272-7358(87)90011-0

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science, 7,* 411–426. doi:10.1177/1745691612454303

Sackett, D. L., Strauss, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-based medicine: How to practice and teach EBM* (2nd ed.). New York, NY: Churchill Livingston.

Salovey, P., & Turk, D. C. (1991). Clinical judgment and decision-making. In C. R. Snyder & D. R. Forsyth (Eds.), *Handbook of social and clinical psychology: The health perspective* (pp. 416–437). New York, NY: Pergamon.

Sapyta, J., Riemer, M., & Bickman, L. (2005). Feedback to clinicians: Theory, research, and practice. *Journal of Clinical Psychology, 61,* 145–153. doi:10.1002/jclp.20107

Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: Therapist effects and the contribution of patient severity and risk. *Journal of Consulting and Clinical Psychology, 80,* 535–546. doi:10.1037/a0028898

Sears, D. O., & Whitley, R. E. (1973). Political persuasion. In I. de Sola Pool, W. Schramm, F. W. Frey, N. Maccoby, & E. B. Parker (Eds.), *Handbook of communication* (pp. 253–289). Chicago, IL: Rand-McNally.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes, 53,* 252–266. doi:10.1016/0749-5978(92)90064-E

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78,* 298–311. doi:10.1037/a0019247

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32,* 752–760. doi:10.1037/0003-066X.32.9.752

Snyder, C. R., Shenkel, R. J., & Lowery, C. R. (1977). Acceptance of personality interpretations: The "Barnum effect" and beyond. *Journal of Consulting and Clinical Psychology, 45,* 104–114. doi:10.1037/0022-006X.45.1.104

Snyder, M., & Campbell, B. (1980). Testing hypotheses about other people: The role of the hypothesis. *Personality and Social Psychology Bulletin, 6,* 421–426. doi:10.1177/014616728063015

Spengler, P. M., White, M. J., Ægisdóttir, S., Maugherman, A. S., Anderson, L. A., Cook, R. S., . . . Rush, J. D. (2009). The Meta-Analysis of Clinical Judgment Project: Effects of experience on judgment accuracy. *The Counseling Psychologist, 37,* 350–399. doi:10.1177/0011000006295149

Stein, D. M., & Lambert, M. J. (1984). On the relationship between therapist experience and psychotherapy outcome. *Clinical Psychology Review, 4,* 127–142. doi:10.1016/0272-7358(84)90025-4

Stein, D. M., & Lambert, M. J. (1995). Graduate training in psychotherapy: Are therapy outcomes enhanced? *Journal of Consulting and Clinical Psychology, 63,* 182–196. doi:10.1037/0022-006X.63.2.182

Strasser, J., & Gruber, H. (2004). The role of experience in professional training and development of psychological counsellors. In H. P. A. Boshuizen, R. Bromme, & H. Gruber (Eds.), *Professional learning: Gaps and transitions on the way from novice to expert* (pp. 11–28). Dordrecht, The Netherlands: Kluwer Academic. doi:10.1007/1-4020-2094-5_2

Strohmer, D. C., Shivy, V. A., & Chiodo, A. L. (1990). Information processing strategies in counselor hypothesis testing: The role of selective memory and expectancy. *Journal of Counseling Psychology, 37,* 465–472. doi:10.1037/0022-0167.37.4.465

Strupp, H. H., & Hadley, S. W. (1979). Specific vs. nonspecific factors in psychotherapy: A controlled study of outcome. *Archives of General Psychiatry, 36,* 1125–1136. doi:10.1001/archpsyc.1979.01780100095009

Tracey, T. J. G., & Rounds, J. (1999). Inference and attribution errors in test interpretation. In R. K. Goodyear & J. W. Lichtenberg (Eds.), *Test interpretation: Integrating science and practice* (pp. 113–131). Boston, MA: Allyn & Bacon.

Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem-solving skill in the social sciences. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 17, pp. 165–213). New York, NY: Academic Press.

Voss, J. F., & Post, T. A. (1988). On the solving of ill-structured problems. In M. T. H. Chi R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. 261–285). Hillsdale, NJ: Erlbaum.

Voss, J. F., Tyler, S. W., & Yengo, L. A. (1983). Individual differences in the solving of social science problems. In R. Dillon & R. Schmeck (Eds.), *Individual differences in cognition* (Vol. 1, pp. 205–232). New York, NY: Academic Press.

Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports, 110,* 639–644.

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61,* 620–630. doi:10.1037/0022-006X.61.4.620

Wampold, B. E. (2001a). Contextualizing psychotherapy as a healing practice: Culture, history, and methods. *Applied and Preventative Psychology, 10,* 69–86.

Wampold, B. E. (2001b). *The great psychotherapy debate: Models, methods, and findings.* Mahwah, NJ: Erlbaum.

Wampold, B. E., & Brown, G. S. (2005). Estimating therapist variability: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology, 73,* 914–923. doi:10.1037/0022-006X.73.5.914

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bonafide psychotherapies: Empirically, "all must have prizes". *Psychological Bulletin, 122,* 203–215. doi:10.1037/0033-2909.122.3.203

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78,* 200–211. doi:10.1037/a0018912

Wedding, D., & Faust, D. (1989). Clinical judgment and decision making in neuropsychology. *Archives of Clinical Neuropsychology, 4,* 233–265. doi:10.1016/0887-6177(89)90016-4

Wierzbicki, M. (1993). *Issues in clinical psychology: Subjective versus objective approaches.* Boston, MA: Allyn & Bacon.

Williams, E. F., Dunning, D., & Kruger, J. (2013). The hobgoblin of consistency: Algorithmic judgment strategies underlie inflated self-assessment of performance. *Journal of Personality and Social Psychology, 104,* 976–994. doi:10.1037/a0032416

Witteman, C. L. M., & Van den Bercken, J. H. L. (2007). Intermediate effects in psychodiagnostic classification. *European Journal of Psychological Assessment, 23,* 56–61. doi:10.1027/1015-5759.23.1.56

Witteman, C. L. M., Weiss, D. J., & Metzmacher, M. (2012). Assessing diagnostic expertise of counselors using the Cochran-Weiss-Shanteau (CWS) Index. *Journal of Counseling and Development, 90,* 30–34. doi:10.1111/j.1556-6676.2012.00005.x

Law and Psychiatry

WILLIAM H. REID, MD, MPH

Treating Clinicians and Expert Testimony

Civil and criminal attorneys often refer their clients to psychiatrists or counselors for "treatment" in anticipation of a later report or expert testimony. They may also seek out professionals who have treated the client earlier to help with the legal case. This month's column highlights the inadvisability—and sometimes impropriety—of a *treating* clinician's becoming a *forensic* consultant or expert in the same case.

TREATER–EXPERT CONFLICT

Although a clinician may report the "facts" of his or her experience with the patient, given appropriate permission and disclosures, problems arise when a treating professional either 1) fails to disclose to the court the possible conflict of interest involved in having a current or past treatment relationship* (and thus being obligated to protect the patient's interest) *or* 2) offers professional opinions about the patient or legal case (in court, professional "opinions" can only be offered by expert witnesses). It is usually inappropriate, and a disservice to the court, for a doctor or therapist to assume the dual role of treater and expert witness.

First, a *treatment* relationship creates a professional, ethical, and legal (or "fiduciary"†) obligation to act in the patient's best interest both during and after the treatment relationship. Since *forensic* reports and testimony require objectivity regardless of the patient's wishes or needs, an inherent conflict is created. This conflict is recognized in the ethical guidelines of both the American Psychological Association and the American Psychiatric Association.

Second, the clinician who testifies regarding a current or past patient knows (or should know) that he or she is required to act in the patient's interest, and may even have a personal affinity for the patient's viewpoint. This creates a danger of *intentional bias*.

Third, separate from the clinician's conscious awareness, the obligation to "do no harm" is so keenly felt by ethical practitioners that there is a danger of *unintended bias* toward the patient.

Fourth, when a treating clinician anticipates reporting to a third party (such as a lawyer, court, or insurance company), professional ethics require that this be discussed with the patient as early as is feasible. The awareness of potential disclosure affects the patient's revelations to the clinician, and thus the validity of any report or testimony.

Fifth, the clinician's role and training are not forensic. Even when they know litigation is involved, treating clinicians rarely corroborate patient or case information to the same extent as forensic consultants, and usually have not completely disclosed to the patient any responsibility they have to report to a lawyer or court. Further, they often have a limited or simplistic view of the legal case and the rules that govern it, making them vulnerable to forensic misunderstanding and, at worst, manipulation by the attorney.

SPECIAL CIRCUMSTANCES

Civil Commitment Cases

There are a few administrative and legal matters in which treating clinicians may ethically offer professional opinions. In civil commitment cases, one may speak to the need for involuntary hospitalization, but the abridgement of the patient's freedom has a treatment purpose and is thus in his or her best interest.

*Note that a clinician-patient relationship is formed whenever counseling or treatment is undertaken or anticipated, regardless of who referred the patient.

†A fiduciary is a person or organization who is legally required to put the patient's or client's interests ahead of his/its own. Other examples include bank trustees and court-appointed guardians.

Dr. Reid is a forensic psychiatrist in Horseshoe Bay, Texas, and a past president of the American Academy of Psychiatry and the Law. His most recent book, *The Treatment of Psychiatric Disorders, 3rd Edition*, will soon be followed by *Legal Issues for Psychotherapists*. His web page, Psychiatry and Law Updates, may be found at
<www.reidpsychiatry.com>

This column contains general clinical and clinical-forensic opinions which should not be construed as applying to any specific case, nor as any form of legal advice.

Reports to Insurance Companies, Utilization Reviewers, or Disability Agencies

Reports to insurance companies, utilization reviewers, or disability agencies create a bit more conflict. It is important that the patient understand and accept the clinician's need to report or divulge information. It is just as important that the clinician be as accurate and objective as possible, and be aware of the various temptations to, for example, cast the patient's behavior and diagnosis in a light that favors reimbursement, or inappropriately emphasize symptoms that support a disability claim. We expect our patients to be honest with us; it is wrong to model dishonesty in our work, even when purporting to help our patients.

Forensic or Correctional Institutions

Mental health professionals who work in forensic or correctional institutions are in a special situation, but are not immune from ethical and fiduciary issues.

> A psychiatrist in a state forensic hospital treated a defendant who was incompetent to stand trial. When he became competent, the psychiatrist was subpoenaed to testify in a trial that could have resulted in the defendant's imprisonment or execution. Since the defendant was incompetent when referred to the hospital, and thus was arguably not able to understand any disclosure or disclaimer a clinician might make to him, the psychiatrist was concerned that testifying might be unethical.

The psychiatrist was right to be concerned. Treating mental health professionals cannot change their ethical requirements just because the hospital has a special name like "forensic" or "prison." Although information concerning what he actually saw or heard ("fact" information) may be elicited from the mental health professional if the court allows it, he is not obligated to offer opinions (an "expert" act) and probably should not do so.

A better course for hospitals that are routinely required to provide forensic reports and expert witnesses is to employ a separate professional for forensic assessment, reports, and testimony. Such professionals avoid forming a clinician-patient relationship (e.g., do not prescribe, treat, or give clinical advice). They should be qualified to do forensic evaluations, clearly identify themselves and their roles to the "evaluee" (n.b., not "patient") whenever the person is seen, and assess the evaluee's competence to understand the disclosure.

Rural Settings

I am often asked about rural settings that have forensic needs but only one mental health professional qualified to offer expert opinions to a court. Although most communities have at least one doctoral-level mental health professional near enough to meet clinical needs, it may be difficult to find another one who meets both criteria for forensic matters: absence of past or present clinical relationship and qualification to work as a forensic expert. Of the two requirements, the absence of current or past relationship is arguably the more important. In most cases, the court's primary need is for an objective clinician, not necessarily one who understands legal nuances. The dual treater-testifier role can almost always be avoided by finding a nontreating professional a few miles away.

Child Custody Evaluations

A recent survey by our office confirmed that child custody evaluations are particularly vulnerable to bias and inappropriate reports or testimony. The general principles of forensic work are highlighted in the cauldron of divorce, acrimony, the child's needs, and sometimes manipulation and intrigue. My opinion is simple: *Treating clinicians, especially parents' therapists, should not offer clinical or legal opinions in custody matters.* They should not ignore subpoenas to provide *factual* information, but should refrain from offering opinions about custody. The mother's, father's, and child's therapists must be as free as possible to provide treatment, and their patients must feel as little fear or implication that the therapist will help or hinder their custody wishes as is possible in such an emotionally laden setting. A separate professional, well qualified in child psychology or psychiatry and child custody settings, should see *all* parties (never just one parent, for example) in an evaluation, not a "helping," format. Protecting the interests of the child requires reviewing the records of other professionals' therapy sessions; however, this must be done with the knowledge that treating clinicians' notes are often biased toward one parent or the other.

FORENSIC QUALIFICATIONS

Placing clinicians into forensic roles when they do not have considerable, relevant forensic and clinical experience can, of course, be problematic. The forensic expert should usually have a terminal degree in his or her field (MSW, PhD, MD with psychiatric training) and be generally familiar with the legal issue at hand.

> A man was convicted of murdering his ex-wife and sent to prison. The killing occurred in his home. The woman's family sued to recover damages from his homeowner's insurance by alleging that his act arose out of mental illness and was thus not really a "murder." The perpetrator had no history of mental ill-

ness and had not pursued any defense of incompetence or non-responsibility. Videotapes of him being interviewed by police within an hour of the killing showed no indication of significant mental impairment, nor did psychiatric interviews for the defendant in the civil lawsuit.

Although the plaintiff's attorneys could find no psychiatrist or clinical psychologist who would say the perpetrator was legally insane at the time of the killing, they retained a local family counselor The counselor, while perhaps a good therapist, had no forensic experience and did not have a license to diagnose or independently treat severe mental illness in that state. His report nevertheless contained diagnoses and sweeping statements to the effect that the very fact that the perpetrator killed someone defined him as legally insane.

The report was easily rebutted by a forensic psychiatrist testifying for the insurance company, and the family counselor was somewhat embarassed by the whole affair.

IS THERE A NEED FOR SPECIAL ETHICAL GUIDELINES IN FORENSIC MATTERS?

The extent to which forensic mental health professionals are subject to clinical ethics (especially in criminal cases) is a matter of some debate. Absence of a clinician-patient relationship deals with the issue of fiduciary duty, but does not exempt us from the ethics of our profession. Some scholars, notably Dr. Paul Appelbaum, have described forensic roles and settings which, they believe, deserve special ethical guidelines. Such exceptions, while not allowing the forensic psychiatrist or psychologist to shed completely the mantle of "clinician," do let him or her carry out legitimate obligations to the court.

THE LAST WORD

Once the role of "treater" has begun, your duty to the patient's interest is clear, compelling, and (barring protecting someone from imminent harm) permanent. It is very difficult to serve the patient and the court at the same time.

EDITORIAL

A Cautionary Lesson from Simulated Patients

Gerald M. Rosen, PhD, and William R. Phillips, MD, MPH

J Am Acad Psychiatry Law 32:132–3, 2004

Ekman and O'Sullivan[1] once asked, "Who can catch a liar?"—and they demonstrated that it was not mental health clinicians. As observed by Slovenko, "A good poker player probably knows better than a mental health professional whether or not a person is lying. A psychiatrist is a doctor, not a lie-detector" (Ref. 2, p 122). Six actors recently provided a dramatic demonstration of these concerns when they feigned the symptoms of Posttraumatic Stress Disorder (PTSD) at a clinic specializing in the assessment and treatment of that disorder; all were accepted as genuine.[3]

An extensive body of literature, heretofore ignored by mental health and medicolegal experts, further documents the inability of health professionals to identify individuals who feign disorders. These studies test physicians with "simulated patients"—normal persons trained to mimic the typical signs and symptoms of common disorders. This use of pseudopatients has its origins in the 1960s, when standardized clinical vignettes were developed to teach and test clinical skills in medical trainees.[4] Over time, the method was extended to assess physicians in community practice and health organizations.[5]

In a search of the medical literature, we identified 12 studies in which (1) normal persons presented significant clinical complaints as simulated patients (SPs), and (2) physicians were provided with a mechanism to report patients suspected to be simulators.[6–17] In all 12 studies, doctors detected simulators at low rates, ranging from 0 percent to 25 percent. Most studies simply reported the percentage of simulators whom physicians correctly identified, but Gordon *et al.*[8] provided additional and important data. These authors recruited 54 interns and trained six SPs to feign one of three clinical problems (urinary frequency, cough, and headache). A total of 233 SP cases resulted, of which only 22 (9.4%) were correctly identified by physicians as "definitely" not genuine. When the standard of judgment or level of confidence was reduced from "definite" to "probable," the number of correctly identified simulators increased to 56 (24.0%). Physicians also had 477 consultations with genuine patients and incorrectly labeled 10 (2.3%) as simulators when making "definite" judgments. When the standard of confidence was lowered to "probable," the rate of false positives increased; 45 (9.4%) genuine patients were misidentified as simulators.

It might be argued that studies using SPs overestimate the likelihood that physicians can be fooled, since clinicians are denied the additional information that may result from repeated visits and an ongoing relationship. However, no studies demonstrate that these factors improve physicians' detection of feigned disorders. Further, malingerers can be consistent when misreporting,[18] and lie detection is not necessarily more accurate in ongoing relationships.[19]

Findings on simulated patients and the general literature on lie detection demonstrate that clinicians are not skilled in judging the credibility of their patients. In the context of a physician-patient relationship, in which a working alliance must be developed, there are good reasons to accept subjective complaints at face value. In the context of legal proceedings, however, physicians should be more circumspect. Testimony should be based on objective findings and the awareness that we all can be fooled. Treating physicians bear special responsibility, since their testimony can create "echo attributions,"

Dr. Rosen is Clinical Professor, Department of Psychology, and Dr. Phillips is Clinical Professor, Department of Family Medicine, University of Washington, Seattle, WA. Address correspondence to: Gerald M. Rosen, PhD, 2825 Eastlake Center, Suite 205, Seattle, WA 98102. E-mail: grosen@u.washington.edu

wherein a false perception of validity attaches to a message delivered by a prestigious source.[20] The problem can be illustrated by the patient who reports a subjective symptom like "nightmares," after which the doctor testifies in court that "the patient suffers from nightmares." Such a declaration, untempered by the evidence from SP studies, creates a false sense of certainty. Clinicians who rely on their patient's reports are advised to state the subjective and objective findings and offer their professional assessment. When questioned about the actual occurrence of subjective symptoms, or the truthfulness of a patient's report, the wise clinician would do well to be less than certain.

References

1. Ekman P, O'Sullivan M: Who can catch a liar? Am Psychol 46: 913–20, 1991
2. Slovenko R: Psychiatry in Law/Law in Psychiatry. New York: Brunner-Routledge, 2002
3. Hickling EJ, Blanchard EB, Mundy E, *et al*: Detection of malingered MVA related posttraumatic stress disorder: an investigation of the ability to detect professional actors by experienced clinicians, psychological tests, and psychophysiological assessment. J Forensic Psychol Pract 2:33–54, 2002
4. Barrows HS: Simulated Patients: The Development and Use of a New Tool in Medical Education. Springfield, IL: Charles C Thomas, 1971
5. Beullens J, Rethans JJ, Goedhuys J, *et al*: The use of standardized patients in research in general practice. Fam Pract 14:58–62, 1997
6. Burri A, McCaughan K, Barrows H: The feasibility of using the simulated patient as a means to evaluate clinical competence of practicing physicians in a community (a pilot project). Annu Conf Res Med Ed 15:295–9, 1976
7. Carney PA, Eliassen S, Wolford GL, *et al*: How physician communication influences recognition of depression in primary care. J Fam Pract 48:958–64, 1999
8. Gordon J, Sanson-Fisher R, Saunders NA: Identification of simulated patients by interns in a casualty setting. Med Educ 22: 533–8, 1988
9. Kopelow ML, Schnabl GK, Hassard TH, *et al*: Assessment of performance in the office setting with standardized patients: assessing practicing physicians in two settings using standardized patients. Acad Med 67(suppl):S19–21, 1992
10. McLeod PJ, Tamblyn RM, Gayton D, *et al*: Use of standardized patients to assess between-physician variations in resource utilization. JAMA 278:1164–8, 1997
11. Norman GR, Neufeld VR, Walsh A, *et al*: Measuring physicians' performances by using simulated patients. J Med Educ 60:925–34, 1985
12. O'Hagen JJ, Davies LJ, Peans RK: The use of simulated patients in the assessment of actual clinical performance in general practice. NZ Med J 99:948–51, 1986
13. Owen A, Winkler R: General practitioners and psychosocial problems: an evaluation using pseudopatients. Med J Aust 2:393–8, 1974
14. Peabody JW, Luck J, Glassman P, *et al*: Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. JAMA 283:1715–22, 2000
15. Rethans J-J, Drop R, Sturmans F, *et al*: A method for introducing standardized (simulated) patients into general practice consultations. Br J Gen Pract 41:94–6, 1991
16. Rethans J-J, Saebu L: Do general practitioners act consistently in real practice when they meet the same patient twice?: examination of intradoctor variation using standardised (simulated) patients, BMJ 314:1170–3, 1997
17. Woodward CA, McConvey GA, Neufeld V, *et al*: Measurement of physician performance by standardized patients. Med Care 23:1019–27, 1985
18. Wetter MW, Deitsch SE: Faking specific disorders and temporal response consistency on the MMPI-2. Psychol Assess 8:39–47, 1996
19. Swann WB Jr, Silvera DH, Proske CU: On "knowing your partner": dangerous illusions in the age of AIDS? Pers Relationships 2:173–86, 1997
20. Rosen GM: Malingering and the PTSD Database, in Posttraumatic Stress Disorder: Issues and Controversies. Edited by Rosen GM. Chichester, UK: John Wiley & Sons, in press

aps
ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science in the Public Interest
11(3) 89–121

DOI: 10.1177/1529100610390861
http://psi.sagepub.com
SAGE

Pitfalls and Opportunities in Nonverbal and Verbal Lie Detection

Aldert Vrij[1], Pär Anders Granhag[2], and Stephen Porter[3]
[1]University of Portsmouth, [2]University of Gothenburg, [3]University of British Columbia

Summary

The question of whether discernible differences exist between liars and truth tellers has interested professional lie detectors and laypersons for centuries. In this article we discuss whether people can detect lies when observing someone's nonverbal behavior or analyzing someone's speech. An article about detecting lies by observing nonverbal and verbal cues is overdue. Scientific journals regularly publish overviews of research articles regarding nonverbal and verbal cues to deception, but they offer no explicit guidance about what lie detectors should do and should avoid doing to catch liars. We will present such guidance in the present article.

The article consists of two parts. The first section focuses on pitfalls to avoid and outlines the major factors that lead to failures in catching liars. Sixteen reasons are clustered into three categories: (a) a lack of motivation to detect lies (because accepting a fabrication might sometimes be more tolerable or pleasant than understanding the truth), (b) difficulties associated with lie detection, and (c) common errors made by lie detectors. We will argue that the absence of nonverbal and verbal cues uniquely related to deceit (akin Pinocchio's growing nose), the existence of typically small differences between truth tellers and liars, and the fact that liars actively try to appear credible contribute to making lie detection a difficult task. Other factors that add to difficulty is that lies are often embedded in truths, that lie detectors often do not receive adequate feedback about their judgments and therefore cannot learn from their mistakes, and that some methods to detect lies violate conversation rules and are therefore difficult to apply in real life. The final factor to be discussed in this category is that some people are just very good liars.

The common errors lie detectors make that we have identified are examining the wrong cues (in part, because professionals are taught these wrong cues); placing too great an emphasis on nonverbal cues (in part, because training encourages such emphasis); tending to too-readily interpret certain behaviors, particularly signs of nervousness, as diagnostic of deception; placing too great an emphasis on simplistic rules of thumb; and neglecting inter- and intrapersonal differences. We also discuss two final errors: that many interview strategies advocated by police manuals can impair lie detection, and that professionals tend to overestimate their ability to detect deceit.

The second section of this article discusses opportunities for maximizing one's chances of detecting lies and elaborates strategies for improving one's lie-detection skills. Within this section, we first provide five recommendations for avoiding the common errors in detecting lies that we identified earlier in the article. Next, we discuss a relatively recent wave of innovative lie-detection research that goes one step further and introduces novel interview styles aimed at eliciting and enhancing verbal and nonverbal differences between liars and truth tellers by exploiting their different psychological states. In this part of the article, we encourage lie detectors to use an information-gathering approach rather than an accusatory approach and to ask liars questions that they have not anticipated. We also encourage lie detectors to ask temporal questions—questions related to the particular time the interviewee claims to have been at a certain location—when a scripted answer (e.g., "I went to the gym") is expected. For attempts to detect lying about opinions, we introduce the devil's advocate approach, in which investigators first ask interviewees to argue in favor of their personal view and then ask them to argue against their personal view. The technique is based on the principle that it is easier for people to come up with arguments in favor than against their personal view. For situations in which investigators possess potentially incriminating information about a suspect, the "strategic use of evidence" technique is introduced. In this technique, interviewees are encouraged to discuss their activities, including those related to the incriminating information, while being unaware that the interviewer possesses this information. The final technique we discuss is the "imposing cognitive load" approach. Here, the assumption is that lying is often more difficult than truth telling. Investigators could increase the differences in cognitive load that truth

Corresponding Author:
Aldert Vrij, University of Portsmouth, Psychology Department, King Henry Building, King Henry I Street, Portsmouth, United Kingdom PO1 2DY
E-mail: aldert.vrij@port.ac.uk

tellers and liars experience by introducing mentally taxing interventions that impose additional cognitive demand. If people normally require more cognitive resources to lie than to tell the truth, they will have fewer cognitive resources left over to address these mentally taxing interventions when lying than when truth telling. We discuss two ways to impose cognitive load on interviewees during interviews: asking them to tell their stories in reverse order and asking them to maintain eye contact with the interviewer.

We conclude the article by outlining future research directions. We argue that research is needed that examines (a) the differences between truth tellers and liars when they discuss their future activities (intentions) rather than their past activities, (b) lies told by actual suspects in high-stakes situations rather than by university students in laboratory settings, and (c) lies told by a group of suspects (networks) rather than individuals. An additional line of fruitful and important research is to examine the strategies used by truth tellers and liars when they are interviewed. As we will argue in the present article, effective lie-detection interview techniques take advantage of the distinctive psychological processes of truth tellers and liars, and obtaining insight into these processes is thus vital for developing effective lie-detection interview tools.

Introduction

"Deception entered Western thought in a telling guise when the author of *Genesis* placed a serpent in the Garden of Eden. By lying, the serpent enticed Eve into committing the original sin" (C.F. Bond & DePaulo, 2006, p. 214). Lying has always posed a moral problem. For example, St. Augustine believed that every lie is a sin, and Aristotle and Kant expressed similar views. In contrast, Machiavelli highly praised deceit in the service of self (Bok, 1989; C.F. Bond & DePaulo). The nature of lying is two-pronged, and how we feel about deception depends on the reason for telling the lie (Seiter, Bruschke, & Bai, 2002). Most lies are told for psychological reasons, and people do not feel bad about telling these kinds of lies. We do not relish having to express all of our thoughts (e.g., "I find that woman more attractive than my own partner.") and thus, we would rather lie. Instead of always showing our true selves, we prefer to censor ourselves so that we are perceived by others in a positive light. We tell psychological lies for a number of reasons: to protect ourselves, to avoid tension and conflict in social interactions, and to minimize hurt feelings and ill will (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996).

However, sometimes the situation is different, such as when people really would like to know the truth; these situations can arise during activities such as watching the evening news or interviewing a candidate for employment. For example, a viewer may want to know whether a politician's denial of involvement in a bribery scandal is really the truth; a teacher may want to know whether a student has cheated during the exam he or she aced; a mother may want to know whether her daughter really has finished her homework; the potential buyer of a used car wants to know whether the vehicle is really as good as the salesperson says; an interviewer may want to know whether the candidate is indeed as capable as he or she claims; a customs officer may want to know whether the traveler really has nothing to declare; an airport security officer wants to know whether the passenger really has no harmful intent when entering the aircraft; and a police detective wants to know whether a suspect's alibi is reliable. Successfully detecting lies in situations such as these would benefit individuals and the society as a whole.

For centuries, the question of whether discernable differences exist between liars and truth tellers has interested practitioners and laypersons (Trovillo, 1939). Throughout history, people have assumed that lying is accompanied by physiological activity in the liar's body. For example, in 1000 B.C., the Chinese forced suspected liars to chew rice powder and then spit it out. If the resultant powder was dry, then the person was judged to have been lying (Kleinmuntz & Szucko, 1984). There was a physiological basis for this assumption. Liars were assumed to fear being caught, and fear is associated with decreased salivation and a dry mouth (Ford, 2006). Nowadays, technology is used to measure physiological (and neurological) reactions—particularly the polygraph; voice-stress analyzers; electroencephalograms (EEG); and most recently, functional magnetic resonance imaging (fMRI). The promotion of such tools can be aggressive. For example, companies have begun to offer fMRI deception-detection services to investigators. Two companies—Cephos Corporation in Massachusetts and No Lie MRI, Inc. in California—claim to know with at least 90% accuracy whether a subject is telling the truth (Stix, 2008). However, a very small number of published studies have examined brain function during deception, and such claims lack strong empirical foundation (Greely & Illes, 2007; Porter & ten Brinke, 2010; Spence, 2008; Wolpe, Foster, & Langleben, 2005). Specifically, Spence (2008) points to problems with replication, large individual brain differences, and no clear brain regions associated with truth telling. Also, brain activity when lying varies depending on the situation. Ganis, Kosslyn, Stose, Thompson, and Yurgelun-Todd (2003) found that telling spontaneous lies corresponds to activation in different brain areas than does telling rehearsed lies; feeling strongly about the topic under investigation and the negative consequences of getting caught also corresponds to different brain activity than feeling less strong.

In this article, we neither discuss physiological or neurological cues to deceit nor focus on lie-detection tools that use equipment. Rather, we focus on an individual's overt nonverbal behavior or speech that human perceivers can discern without the aid of equipment. Further, we address whether people can detect lies when observing someone's nonverbal behavior or when analyzing someone's speech. This technique—observation—is the most common form of lie detection; in many situations, technologies that are used to measure physiological or neurological cues are unavailable or are not possible to implement.

In our view, research on lie detection through observations of nonverbal and verbal cues is overdue. Scientific journals

regularly publish overviews of research articles regarding nonverbal and verbal cues of deception (for recent examples, see DePaulo et al., 2003; Masip, Sporer, Garrido, & Herrero, 2005; Sporer & Schwandt, 2006, 2007; Vrij, 2005). These meta-analyses provide valuable information about how liars behave and the stories they tell, but they offer no explicit guidance about what lie detectors should do and avoid doing in order to detect deception.

This article consists of two sections. The first section focuses on pitfalls to avoid and outlines the major factors that lead to failures in detecting liars: We cluster 16 reasons into three categories (Vrij, 2007, 2008a): (a) a lack of motivation to detect lies, (b) difficulties associated with lie detection, and (c) common errors made by lie detectors. Discussing pitfalls is important because it provides insight into how lie detectors can improve their performance (e.g., by recognizing common biases and by avoiding common judgment errors). The second section of this article discusses opportunities for maximizing one's chances of detecting lies and elaborates on strategies for improving one's lie-detection skills. In this section, we first provide five recommendations for avoiding common errors in detecting lies. These recommendations are firmly based in a rich body of psychological research over the past few decades. Next, we discuss a relatively recent wave of innovative lie-detection research that goes one step further by introducing novel interview styles aimed at eliciting and enhancing verbal and nonverbal differences between liars and truth tellers by exploiting their different psychological states. The recommendations are relevant in varied walks of life, from the individual level (e.g., "Is my partner really working late to meet a deadline?") to the societal level (e.g., "Can we trust this informant when he claims that he can disclose information about an active terrorist cell in London?").

Before we discuss the common pitfalls associated with lie detection, three issues merit attention: (a) a definition of *deception*, (b) the underlying premises of verbal and nonverbal cues to deception and its detection, and (c) research methods used in deception research.

Defining *deception* is not a straightforward task. Deception has been studied through the lens of varied disciplines, including psychiatry, linguistics, and philosophy; and accordingly, diverse definitions have been offered (Granhag & Strömwall, 2004). In the present context, we deem Vrij's (2008a, p. 15) definition of *deception* to be sufficient: "*a successful or unsuccessful attempt, without forewarning, to create in another a belief which the communicator considers to be untrue.*" It is important to note that lying is an intentional act and that misremembering is not the same as lying.

Researchers have proposed different theoretical approaches to predict which verbal and nonverbal cues to deception may occur, particularly Ekman and Friesen's (1969) leakage and deception cues approach; Zuckerman, DePaulo, and Rosenthal's (1981) multifactor model; Ekman's (1985/2001) emotion approach; Buller and Burgoon's (1996) interpersonal deception theory; and DePaulo's self-presentational perspective (DePaulo, 1992; DePaulo et al., 2003). These approaches have three elements in common that have influenced verbal and nonverbal lie detection: the notion that, compared with truth tellers, liars (a) may experience stronger emotions (particularly fear, as a result of detection apprehension), (b) may experience higher levels of cognitive load, and (c) are inclined to use more and different strategies to make a convincing impression on others.

Traditionally, verbal and nonverbal lie detection has focused on the difference in emotions that liars and truth tellers experience. Ekman's (1985/2001) analysis of microexpressions is a prime example, but also lie-detection techniques promoted in police manuals are primarily based on the notion that liars are more concerned and nervous than truth tellers (Vrij & Granhag, 2007). The approach has limitations. First, experiencing emotions is not the sole domain of liars: Truth tellers can experience the same emotions, particularly if they know that they are scrutinized and/or are afraid of not being believed (e.g., see our later discussion of the Othello error). If emotional displays or cues of nervousness per se do not reliably distinguish between truth tellers and liars, the next step is to ask questions that will elicit such cues in liars but not in truth tellers or, alternatively, that will enhance such cues more in liars than in truth tellers. No such questioning technique exists to date, and it is doubtful that it can ever be developed (National Research Council, 2003). For the latter reason, in more recent lie-detection studies, researchers have concentrated on cognitive load. The premise here is that lying is mentally more taxing than truth telling. This approach shares one limitation with the emotion approach. Cues of cognitive load are not the sole domain of liars either; truth tellers also may have to think hard, and therefore they may display cues of being mentally taxed. However, unlike the emotion approach, interview protocols that elicit and enhance cues of cognitive load more in liars than in truth tellers can be developed, making it possible to discriminate between the two. We elaborate on this concept later in the "Exploiting the Differential Mental Processes of Truth Tellers and Liars" section. The same section also discusses another strain of recent lie-detection research that aims to exploit the fact that liars use more and different strategies to avoid detection than do truth tellers. In sum, in verbal and nonverbal lie detection, the emphasis has moved in recent years from emotion-based lie-detection techniques to cognitive-load lie-detection techniques that focus on liars' and truth tellers' different psychological states and take their differential strategies into account.

We base our analysis of pitfalls and opportunities in nonverbal and verbal lie detection on scientific research. In studies in which researchers have examined nonverbal and verbal cues to deception, trained raters watch video footage or analyze transcripts of such footage of truth tellers and liars. They analyze with particular coding systems the frequency of occurrence or duration of various nonverbal and verbal cues displayed by truth tellers and liars (e.g., all sorts of movements, eye contact, smiles, pauses, amount of detail, type of detail, contradictions) and compare the truthful and deceptive responses. There are two types of studies—those conducted in the field

and those conducted in the laboratory. In real-life studies, typically called "field studies," video footage of real-life settings, such as police–suspect interviews, is analyzed (Mann, Vrij, & Bull, 2002). In laboratory studies, video footage and/or transcripts of participants who were instructed by researchers to tell the truth or lie for the purpose of the experiment are analyzed. Field studies probably have greater appeal because they are realistic. However, conducting field studies is problematic, particularly in establishing the *ground truth*—researchers can analyze only the responses known to be true or false. To establish this ground truth satisfactorily, independent case facts, such as medical evidence, material evidence, DNA evidence, or reliable eyewitnesses, are needed. Unfortunately, such facts are often unavailable. In laboratory studies, researchers (a) ask participants (mostly college students) to tell the truth or lie and (b) measure their nonverbal and verbal responses during both activities. In the studies published to date, participants have told the truth or lied about many different topics—a film they had just seen, possession of a certain object in their pocket, their involvement in the disappearance of some money, the number of dots that appeared on a screen, their feelings about certain people, or their opinions about controversial issues. More recently, researchers have introduced scenarios that better reflect forensic real-life situations. In a study by Hartwig, Granhag, Strömwall, and Kronkvist (2006), participants were sent to a shop to buy a product (truth tellers) or steal a wallet (liars) and were interviewed about the alleged shop visit. In a study by Vrij, Leal, Mann, and Granhag (in press), participants were sent to receive a package at a certain location and deliver it somewhere else and were then interviewed about this mission (liars had to hide the details of what they did). In study by Strömwall, Granhag, and Jonsson (2003), participants (a) were sent to a restaurant to have lunch (truth tellers) or (b) committed a mock crime (liars) and were asked to pretend that they had had lunch in a restaurant. And in a study by Vrij, Granhag, Mann, and Leal (in press), passengers at an international airport were asked to tell the truth or lie about their forthcoming trip. The advantage of laboratory studies is that researchers can establish the ground truth. However, laboratory studies have limitations. In such studies, participants do not choose to lie, but rather they are instructed to do so by the experimenter, meaning that lying is condoned. Another restriction is that the stakes (negative consequences of being caught or positive consequences of being believed) are never really high (Ekman, 1985/2001; Malone & DePaulo, 2001; Miller & Stiff, 1993). To raise the stakes in laboratory experiments, participants have been offered money if they succeed in lying (Vrij, Akehurst, Soukara, & Bull, 2002; Vrij, Edward, & Bull, 2001). In other studies, participants are told that they will be observed by their peers, who will judge their sincerity (DePaulo, Stone, & Lassiter, 1985), or told that being a good liar is an important indicator of being successful in a future career (DePaulo, Lanier, & Davis, 1983). Such studies provide useful examples of how people behave when they lie in daily life, because most of the lies people tell are low-stakes lies (DePaulo et al., 1996).

However, suspects in police interviews, smugglers at airports, corrupt politicians in conversations with suspicious journalists, and husbands who cheat on their wives all tell high-stakes lies. In an attempt to create examples of such lies, some researchers have raised the stakes further in laboratory studies. For example, participants in Frank and Ekman's (1997) experiment were given the opportunity to "steal" US $50. If they could convince the interviewer that they had not taken the money, they could keep all of it. If they took the money and the interviewer judged them as lying, they had to return the US $50 and they would also lose their US $10-per-hour participation fee. Moreover, some participants faced an additional punishment if they were found to be lying. They were told that they would have to sit on a cold, metal chair inside a cramped, darkened room ominously labeled "XXX," where they would have to endure anything from 10 to 40 randomly sequenced 110-decibel starting blasts of white noise over the course of 1 hour.

A study such as the one just mentioned raises ethical concerns. Yet, even apart from this concern, one might argue that the stakes in such a study do not compete with the stakes in some real-life situations. Providing even larger incentives to participants is always possible. For example, participants in Frank and Ekman's (1997) study could have been offered US $500 instead of US $50 if they succeed in convincing the interviewer that they are telling the truth. Introducing severe punishments for those who fail to convince the interviewer that they are telling the truth is, however, not possible, because university ethics committees will not approve such experiments. Also, punishments are never realistic, and participants may be aware of it. Ethical guidelines require researchers to inform participants before participation that they are free to withdraw from the study at any time. Hence, when participants are threatened with having to enter a dark room to face white noise for 1 hour, as in Frank and Ekman's study, they will realize that they are actually free to leave. In other words, it may not be possible to introduce truly high-stakes settings in laboratory experiments, and thus, examining how liars behave in high-stake real-life situations is often the only option (Barrett, 2005; Riggio, 1994).

In a typical lie-detection study, observers (often undergraduate students, but sometimes professionals such as police officers or police detectives) are shown short video fragments of people they do not know who are either telling the truth or lying. The fragments the observers have to judge are typically derived from the studies that have been discussed in the previous paragraph. The observers are asked to indicate after each fragment whether the person (often called the *sender*) was telling the truth or lying. Typically, half of the senders are truth tellers, and half are liars. (The observers are typically not informed what percentage will be truth tellers and liars, because this may result in them deliberately trying to achieve an equal number of truth and lie responses.) In such a study, simply guessing whether the sender spoke the truth or lied would result in correctly classifying 50% of the truths (truth accuracy rate) and 50% of the lies (lie accuracy rate),

resulting in a total accuracy rate (truth and lie accuracy rate combined) of 50%.

In lie-detection studies, observers are typically not given any background information about the senders and their statements, so the only source of information available to them is the senders' nonverbal and verbal behavior. (Exceptions are the "Strategic Use of Evidence" studies, which are discussed later in this article.) Such a situation is not typical of lie-detection in real life. In their study, Park, Levine, McCornack, Morrisson, and Ferrara (2002) asked college students (a) to recall an instance in their life in which they had detected that another person had lied to them and (b) to report how they had discovered the lie. Participants detected less than 2% of the lies by relying exclusively on the liars' nonverbal behavior or speech content at the time the lies were told. More commonly, participants discovered the lies through information from third parties (38%), physical evidence (23%), and confessions (14%). More than 80% of the lies were detected 1 hour or more after they were told, and 40% were detected more than a week later.

Pitfalls in Lie Detection

Lack of motivation to catch liars: The ostrich effect

Lies often remain undetected because people do not attempt to uncover the truth (Ekman, 1985/2001), a phenomenon labeled the *ostrich effect* (Vrij, 2008a). A fabrication might sometimes be more tolerable or pleasant than the truth for the message recipient, rendering ignorance the preferred option. For example, why bother trying to discover whether mendacious compliments about one's body shape, hairstyle, dress sense, or achievements are truthful?

For this reason, the ostrich effect extends to more serious lies, which thus also remain undiscovered. For example, Betty Currie, who was former U.S. President Bill Clinton's secretary, tried to avoid learning details of the relationship between the President and Monica Lewinsky (Vrij, 2008a). Indeed, rather than gain anything from knowing the truth, she would have been put in the difficult position of having to decide what to do with such knowledge. Not knowing what to do when having learned the truth may also be the reason why some people overlook evidence for possible infidelity by their romantic partners, instead remaining in denial (Feldman & Cauffman, 1999). If an individual discovers that his or her partner is having an affair, this discovery could create a difficult situation for the betrayed spouse. For example, there is the risk of the cheating partner leaving the betrayed spouse if confronted with the evidence. If they also have children, the betrayed spouse may feel that marital dissolution is undesirable because of its effect on their children. In such situations, it is worthwhile to engage defense mechanisms such as denial in order to avoid acknowledging the truth. In brief, even though the solution may be worse than the problem, ignorance can be bliss.

Difficulty of lie detection: Absence of Pinocchio's growing nose

In the classic tale *The Adventures of Pinocchio*, Pinocchio's nose grew larger each time he lied, but it was unaltered each time he spoke the truth, so his growing nose was a reliable cue to deceit. The meta-analyses that have been published to date have made clear that there are no nonverbal and verbal cues uniquely related to deceit. In other words, reliable cues to deception akin to Pinocchio's growing nose do not exist (DePaulo et al., 2003; Masip et al., 2005; Sporer & Schwandt, 2006, 2007; Vrij, 2005). The fact that there is no single cue that lie detectors can consistently rely upon makes lie detection inherently difficult.

The meta-analyses further reveal that the majority of the nonverbal and verbal cues that researchers typically examine in deception studies are not related to deception at all. For example, in DePaulo et al.'s (2003) meta-analysis—the most extensive one to date—the researchers investigated 158 cues, of which 118 (75%) showed no association with deception at all (including cues people often associated with lying, such as gaze aversion, postural shifts, pauses, and self-references). Many cues that were found to be to some extent related to deception were often examined sporadically, and it is important for researchers to replicate those cues' diagnostic value before drawing conclusions.

Subtle differences

Another difficulty that lie detectors face is that any behavioral differences between truth tellers and liars are typically small. For example, in DePaulo et al.'s (2003) meta-analysis, 14 of the 50 (28%) cues that had been examined in six or more deception studies revealed a significant association with deception, including liars who provided fewer details and less plausible answers than did truth tellers, and liars who made fewer illustrators (i.e., hand movements that accompany speech and illustrate it) than did truth tellers. However, the average effect size of the relation of the various behaviors with deception was only $d = .25$, which is considered to be a small or modest effect (Cohen, 1977). Because these relationships are modest, police manuals that describe nonverbal and verbal cues of deceit are misleading. Although such manuals often offer brief warnings about the unreliability of cues to deception, those caveats are easily lost in the ensuing detailed and enthusiastic descriptions of how behavior and speech differs between truth tellers and liars (see also Moston, 1992). Those descriptions are sometimes accompanied by photographs demonstrating "truthful forward posture" and "deceptive adaptor behaviors" (Inbau, Reid, Buckley, & Jayne, 2001, pp. 145, 149), thereby suggesting that (a) reliable cues to deception do exist and (b) the differences between truth tellers and liars are substantial and therefore easy to spot. Nevertheless, no scientific research supports these promises: Cues to deception are generally unreliable and faint.

The fact that cues to deception are unreliable and faint aligns with the previous contention that *emotions* and *cognitive load*—two main indicators of deception—can be displayed by both liars and truth tellers. A more promising picture may emerge when interviewers attempt to elicit and enhance cues to deceit. Such studies—discussed later in this article—are scarce and have only recently been conducted; in fact, none of these were published before 2003, the year that DePaulo et al.'s meta-analyses was published.

Countermeasures

A further complication for lie detectors is that liars—particularly those communicating high-stakes lies—often deliberately attempt to appear credible in order to avoid detection; strategies to achieve this goal are called *countermeasures*. A verbal veracity assessment tool widely used by professional lie catchers is statement validity assessment. Statement validity assessments are accepted as evidence in some North American courts (Ruby & Brigham, 1997) and in criminal courts in several West European countries, including Austria, Germany, Sweden, Switzerland, and The Netherlands (Köhnken, 2002, 2004). The statement validity assessment originates from Sweden (Trankell, 1972) and Germany (Arntzen, 1970, 1982, 1983; Undeutsch, 1967, 1982, 1984, 1989) and has been designed to determine the credibility of child witnesses' testimonies in trials for sexual offenses. The core phase of the statement validity assessment is criteria-based content analysis, a list of 19 criteria thought to be more present in truthful accounts than in false ones (including mentioning space and time, replication of conversation, recall of interactions, unexpected complications, and accounts of mental state; for recent statement validity assessment reviews, see Vrij, 2005, 2008a). However, children (and adults) who learn how criteria-based content analysis works can tell stories that sound plausible to experts in using such analysis (Caso, Vrij, Mann, & de Leo, 2006; Joffe & Yuille, 1992; Vrij et al., 2002, Vrij, Akehurst, Soukara, & Bull, 2004b; Vrij, Kneller, & Mann, 2000). Thus, it is possible to become a "sophisticated" liar by using knowledge-based countermeasures.

Liars may further realize that observers pay attention to their behavioral reactions to ascertain their truthfulness. Liars therefore may attempt to control behavior that could betray their lies (Buller & Burgoon, 1996; Burgoon & Buller, 1994; Burgoon, Buller, Floyd, & Grandpre, 1996; Burgoon, Buller, White, Afifi, & Buslig, 1999; Krauss, 1981). In particular, they may avoid exhibiting behaviors they believe will create a dishonest impression, instead trying to display behaviors they believe will make them appear credible (Hocking & Leathers, 1980; Leary & Kowalski, 1990). Gaze aversion and grooming gestures are among the behaviors most widely believed to signal deceptive behavior (see subsequent section), and liars therefore may avoid displaying them. They appear to be successful in avoiding displaying them because gaze aversion and grooming gestures are unrelated to deception (DePaulo et al., 2003).

Embedded lies

Another difficulty that lie detectors face is that lies are often embedded in truths. That is, rather than telling a blatant lie that is entirely untruthful, liars tend more to change specific vital details in an otherwise truthful story. Thus, when a man wants to conceal his illicit activities on, say, a Tuesday night, he could give details of what he really did on Monday night. Thus, most of the statement is truthful, with only a tiny, but vital, lie (e.g., having committed infidelity or murder) embedded (in this case, by omission or denial). Criminal suspects often tell such embedded lies (see Hartwig, Granhag, & Strömwall, 2007; Porter & Yuille, 1995; Strömwall, Granhag, & Landström, 2007). In a similar vein, when examining false identities adopted by criminals, Wang, Chen, and Atabakhsh (2004) found that such fraudsters typically alter only a small portion of their original identity.

Noncriminals who lie often use a similar embedded-lies strategy (DePaulo et al., 2003; Turner, Edgley, & Olmstead, 1975); this has also been demonstrated in experimental research. For example, in Bell and DePaulo's (1996) experiment, art students asked participants their views on a student's work. When the participants disliked the work, they sometimes overstated the specific elements they favored (e.g., the colors used in the painting) and understated what they disliked. In this lie strategy, most of what the participants said was truthful.

Embedded lies hamper the use of statement validity assessments and other verbal veracity assessment tools such as reality monitoring, because they typically examine the quantity and quality of details in a statement (Masip et al., 2005; Vrij, 2005). Lies that are embedded in predominantly truthful statements may be rich in high-quality details typically associated with credible statements, which could give the lie detector the erroneous impression that the statement is truthful. Lie detectors who focus on nonverbal behavior may make a similar mistake if the deceptive element of a liar's story remains unnoticed (e.g., *when* the person went to the gym) and if they overattend the truthful part instead (e.g., *what* the person did at the gym).

No adequate feedback

Another complication in lie detection is that lie detectors often do not receive adequate feedback about their judgments and therefore cannot learn from their mistakes. For feedback to be helpful, it should be provided frequently, reliably, and immediately. Thus, observers should be informed immediately after every interaction with another person whether that person was lying. They could then learn how liars truly behave and what they really say and incorporate such knowledge into improved lie-catching strategies. However, adequate feedback is often unavailable (DePaulo & Kirkendol, 1989). People often never discover that they have been lied to, or such knowledge is gained long after the interaction (Park et al., 2002). In many cases of wrongful conviction, the police and/or judge only find out their credibility assessment errors years or decades after they occur. By the time they learn that they

attributed honesty to a deceptive person or vice versa, it is too late for them to make meaningful changes to their decision-making strategies.

Customs officers also face feedback problems (DePaulo & Pfeifer, 1986). Part of their jobs is to detect smugglers among travelers. From the numerous passengers they decide not to search, they virtually get no feedback at all. Some of them may be smugglers, but once the officers let them pass unsearched, they will almost never find out that they made a mistake. They may not even get adequate feedback from the people they do search. Among the latter may be smugglers whose illegal goods remain undetected despite a search.

Violation of conversation rules

As we show in the "Exploiting the Different Mental Processes of Truth Tellers and Liars" section of this review, the act of lying becomes increasingly difficult when the lie detector asks further probing questions that follow an initial free recall by the target (Toris & DePaulo, 1984; Vrij, 2008a).[1] However, probes in daily-life conversations can violate social norms, being seen as inappropriate, strange, or impolite. Conversation partners may object to requests such as "Could you elaborate on that?" and "Could you repeat what you just said?" and may even end the conversation.

Further, although focusing on a speaker's body movements could benefit the lie detector because the speaker may reveal signs of deceit (DePaulo et al., 2003; Sporer & Schwandt, 2007), such movement scanning would seem strange and inappropriate in daily-life situations. Conversation rules dictate that a listener should look into a speaker's eyes, but the eyes themselves generally do not reveal reliable information about deception (DePaulo et al.; Sporer & Schwandt). Therefore, these conversation rules (i.e., discourage probing questions and maintain eye gaze) can hamper lie detection.

Good liars

A final factor contributing to the complexity of lie detection is that some people are proficient liars. Although surprisingly little research has addressed the features of a good liar, we believe six features may be especially important. The best liars are those individuals (a) whose natural behavior disarms suspicion; (b) who do not find it cognitively difficult to lie; (c) who do not experience emotions such as fear, guilt, or delight when they are lying; (d) who are good actors and who display a seemingly honest demeanor; (e) whose attractiveness may lead to an inference of virtue and honesty; and/or (f) who are "good psychologists."

Regarding the first feature of the proficient deceiver—natural behavior—certain behavioral patterns are associated with honesty and likability. Such behavioral patterns include gaze directed to a conversation partner, smiling, head nodding, leaning forward, direct body orientation, posture mirroring, uncrossed arms, articulate gesturing, moderate speaking rates, a lack of "ums" and "ers," and vocal variety (Buller & Aune, 1988; Ekman, 1985/2001; Tickle-Degnen & Rosenthal, 1990). Some people show such demeanor naturally even when they are lying (e.g., *natural performers*; Ekman, 1997). Natural performers are likely to be good liars because their natural behavior is likely to allay suspicion. Former U.S. President Bill Clinton was blessed with this characteristic, being naturally warm and engaging, and he was able to tell lies that were highly convincing to his audience. To illustrate, he received a standing ovation in response to his assertive denial of having sexual relations with Monica Lewinsky.

Second, effective liars find the act of telling lies to be cognitively unchallenging. They may plan their statements and behavior well in advance of the lie, and this rehearsal probably facilitates the ease of deception. Although it is obvious that liars should prepare a story that sounds plausible, this task is difficult for many people. Vrij and Mann (2001b) described five cases in which people who were suspected of having killed one of their relatives and initially denied having done so. Some of the individuals described made serious mistakes when they planned their stories, which made it easy to discern that they probably were hiding the truth. For example, one individual reported being knocked unconscious for 10 hours, but medical professionals determined that this scenario was impossible. Even liars who are typically well prepared can face unexpected situations that require an explanation. For example, a wife may confront her husband with the telephone number and address of a woman—unknown to her—that she found in his pocket; or a police detective may tell a suspect that he was seen by a witness at the scene of crime directly after it occurred. To lie successfully in these or similar situations, the liar needs a convincing and plausible answer. To spontaneously invent a plausible answer is probably too difficult for many liars, but original thinkers who are mentally creative may be successful in dealing with such immediate cognitive demands.

Third, liars differ in the emotions they experience while communicating a lie. One job applicant may feel guilty or anxious when exaggerating his or her qualifications, whereas another may not. One suspect may experience extreme anxiety when presenting a false alibi, whereas another suspect may remain calm. One student may feel excitement when sensing that the teacher believes his or her excuse for being late (referred to as *duping delight*), whereas another may feel trepidation and guilt. Deceiving others is made easier if the liar does not experience feelings of guilt, fear, or delight, because in that case, no emotional behavior needs to be suppressed. An absence of emotions during deception can be related to (a) an absence of remorse concerning a specific incident (e.g., defrauding a wealthy corporation), (b) being practiced at and feeling confident when lying, or (c) a lack of emotion in general. Psychopathic individuals, for example, have a profound emotional impairment and, accordingly, they experience little fear or remorse, even when telling a high-stakes lie (e.g., Hare, 2006; Porter & Woodworth, 2007). Moreover, people with a powerful imagination and the capacity to believe what they are saying are unlikely to experience guilt or fear. Sometimes such people can come to develop a false belief in their original lies

after the passage of time and are thus not, strictly speaking, lying (e.g., Pickel, 2004).

Fourth, although natural performers and those who experience little cognitive load or emotions when lying make the best liars, those who can effectively mask signs of cognitive load and emotions and concurrently display behavior that appears credible probably also make good liars. This feat requires good acting skills. If such individuals are not natural performers, their lies may raise suspicion, and they should adapt themselves adequately to disarm this suspicion. The sooner they adapt themselves, the more chance they have of successfully disarming suspicion. It is thus crucial to notice suspicion quickly, which requires good decoding skills.

Fifth, elements of physical appearance can promote effective lying. For example, attractiveness and characteristics of faces can lead to inferences of trustworthiness that facilitate the liar's success (e.g., Porter, England, Juodis, ten Brinke, & Wilson, 2008; Porter, Gustaw, & ten Brinke, 2010).

Last, good liars probably also have good insight into another person's thought processes. They have a sense of what other people want to hear and how to convey it persuasively. In that respect, successful lying could be related to emotional intelligence. However, we are not aware of research that has examined this phenomenon (for in-depth discussions of factors that make people good liars, see Vrij, 2008a; Vrij, Granhag, & Mann, in press).

Common Errors Made by Lie Detectors

People fail to catch liars not only because they are unmotivated to catch them or because the lie-detection task is difficult but also because they make systematic errors in the evaluation process. We believe that eight common errors can be identified, which we examine in this section.

Examining the wrong cues

There are widespread beliefs about how people behave and what they say when they lie. Overwhelmingly, both laypersons and professional lie catchers expect liars to act nervously; exhibiting gaze aversion ("liars look away") and displaying grooming gestures ("liars fidget") are among the most popular beliefs (Strömwall, Granhag, & Hartwig, 2004; Taylor & Hick, 2007; The Global Deception Team, 2006; Vrij, 2008a; Vrij, Akehurst, & Knight, 2006).[2] Charles F. Bond conducted an ambitious "beliefs about cues to deception" project that he published under the name *The Global Deception Team*. The team consisted of an international group of researchers from 58 countries, each collecting data from 20 male and 20 female adult residents of his or her country. The participants were asked to write down their response to the question, "How can you tell when people are lying?" The respondents mentioned 103 different beliefs, 9 of which were given by more than 15% of the participants. One cue in particular was prevalent: gaze aversion. People overwhelmingly asserted that liars avert their gaze, and 64% of the participants expressed this belief. Gaze aversion was the most frequently mentioned belief about deception behavior in 51 out of 58 countries. Gaze aversion showed the lowest prevalence in the United Arab Emirates, where it was mentioned by 20% of the participants, making it the eighth most prevalent belief in that country.

Despite their overwhelming endorsement internationally, cues such as gaze aversion and grooming gestures are not reliable cues to deception (DePaulo et al., 2003; Sporer & Schwandt, 2007). Nonetheless, police and other legal professionals are encouraged to use such incorrect cues in detecting lies (Johnson, 2006a, 2006b). For example, in their influential police manual, Inbau et al. (2001) [3] advocated several nonverbal cues as being diagnostic of deception, including avoiding eye contact and grooming gestures, as well as cues such as frequent posture changes, placing hands over mouth or eyes, and lack of illustrators. Of these cues, only a decrease in illustrators has been found empirically to be associated with deception (e.g., DePaulo et al.). Thus, it is not surprising that, in a lie-detection study in which police officers viewed video fragments of suspects telling the truth or lying during their interviews, there was an inverse relation between (a) the endorsement of the lie cues promoted in the Inbau et al. manual and (b) the ability to distinguish suspects' truths and lies (Mann, Vrij, & Bull, 2004). In another study, college students who had been trained in the behavioral cues described by Inbau et al. performed worse on a subsequent lie-detection test than did untrained participants (Kassin & Fong, 1999). Police manuals often advise investigators to pay attention to signs of nervousness when attempting to detect deceit (Vrij & Granhag, 2007), advice that could easily lead to Othello errors (see subsequent section).

How do such false beliefs about lying develop? One likely contributing factor is moral reasoning. The stereotypical but sometimes incorrect view is that lying is "bad" (Backbier, Hoogstraten, & Meerum Terwogt-Kouwenhoven, 1997; Bok, 1989; DePaulo, 2004; DePaulo et al., 1996; Kowalski, Walker, Wilkinson, Queen, & Sharp, 2003; Robinson, 1994; Schweitzer, Hershey, & Bradlow, 2006). C.F. Bond argued that the prominent lying/gaze-aversion myth fits well with this lying-is-bad stereotype (The Global Deception Team, 2006). Because people often avert their gaze when they feel ashamed, they should do so, it is assumed, when engaging in the reprehensible act of lying (DePaulo et al., 2003). Moreover, because lying is bad, liars should feel nervous about the potential for getting caught, and they should exhibit signs of anxiety such as avoiding eye contact, increased fidgeting, and moving around. Because the association of lying and immorality is taught early in life, children as young as 5 to 6 years of age already associate gaze aversion and limb movements with deception (Rotenberg & Sullivan, 2003).

After such stereotypical beliefs are established, they persist for several reasons, including *illusory correlations*, or the perception of associations that do not exist, develop, strengthen, and cause observers to distort their information processing. For example, in Levine, Asada, and Park's (2006) intriguing experiment, observers who were led to believe that someone

was lying subsequently overestimated the amount of gaze aversion that the supposed liar had actually displayed. A second factor is the phenomenon of *confirmation bias*, a tendency to seek information that confirms existing beliefs (Darley & Gross, 1983); in this case, overattending to observations supposedly validates the relation between lying and gaze aversion/nervousness. Third, when people make observations that could disconfirm a false belief, they often disregard or downplay it instead of interpreting the new evidence properly, a phenomenon called *belief perseverance* (C.A. Anderson, Lepper, & Ross, 1980). Researchers have found such phenomena to influence flawed deception detection and evaluation of evidence in legal cases more generally (Porter, Gustaw, & ten Brinke, 2010). Fourth, after observers form a strong opinion that makes sense to them, they often create further reasons to support their view (Strömwall et al., 2004). In fact, an opinion is often strengthened by merely thinking about the topic (Tesser, 1978). Fifth, as previously mentioned, people typically receive inadequate feedback about the validity of their lie-detection judgments, disallowing effective learning and improvements with experience. Ironically, effective learning opportunities may be available to seasoned criminal offenders more so than to legal decision makers. Offenders probably need to lie frequently and effectively in order to succeed in their criminal careers (e.g., Porter & Woodworth, 2007), and they receive frequent and often immediate feedback on whether their attempts to lie are successful. Accordingly, offenders have more correct views about cues to deception than do laypersons and professional lie catchers (Strömwall et al., 2004; Vrij & Semin, 1996). For example, the erroneous stereotypical view that liars increase their movements is not common among offenders (Vrij & Semin).

The combination of how incorrect beliefs originate and why they last could explain the advocacy of such beliefs in many police manuals. These views are based on subjective impressions about verbal and nonverbal behavior displayed by suspects during police interviews rather than on empirical research. Psychological research and theory suggest that these impressions can easily become distorted. Our advice to authors of police manuals, therefore, is to base their writing on science and not subjective impressions.

Overemphasis on nonverbal cues

In a minority of cases, observers rely on speech content when they attempt to detect deceit. This may occur for example with observers who are knowledgeable about the facts that are discussed by the target person. In such cases, the observer typically focuses on the narrative and compares his or her knowledge with the story the target person provides (e.g., Reinhard, Sporer, & Marksteiner, 2009). Second, observers occasionally have access to more than one statement—multiple statements from the same person or statements from different people—and thus focus on the level of consistency between the statements (Granhag & Strömwall, 1999, 2000a, 2000b, 2001; Strömwall & Granhag, 2005, 2007; Strömwall, Granhag, & Jonsson, 2003). Also, observers may rely on verbal cues when they are distinctive, particularly when a statement appears to be against the self-interest of the storyteller (Noller, 1985), such as a confession.

When the observer possesses no factual information, has no statements for comparison, and when the speech content is not distinctive, observers are inclined to pay greater attention to nonverbal behavior than to verbal behavior. For example, Mann et al. (2004) showed 99 British police officers 54 videotaped fragments of police interviews with individuals who were suspected of rape, arson, or murder. The officers were asked to make veracity judgments following each fragment and to report the cues on which they based their decisions. The majority of the cues reported (78%) were nonverbal (also see Porter, Woodworth, & Birt, 2000). Also, when observers notice that someone's nonverbal behavior and speech content are discrepant, they typically rely on the nonverbal channel. For example, a job applicant with a reserved demeanor who claims to be enthusiastic about the job will be perceived as less keen about it than he or she reports (DePaulo, Rosenthal, Eisenstat, Rogers, & Finkelstein, 1978; Hale & Stiff, 1990; Zuckerman, Driver, & Koestner, 1982; Zuckerman, Speigel, DePaulo, & Rosenthal, 1982).

Lie detectors pay so much attention to nonverbal behavior for several reasons. First, people are used to making inferences from nonverbal behavior, including facial expressions. By observing behavior alone, people draw, with reasonable accuracy, many conclusions about other people, including their personality traits (e.g., extraversion, sociability), masculinity, femininity, or sexual orientation. From behavior, it is also possible to discern information about status, dominance, romantic involvement, and relationship potential (Ambady, Bernieri, & Richeson, 2000), and women are able to accurately rate men's interest in infants based only on viewing their faces (Roney, Hanson, Durante, & Maestripieri, 2006). Observing only 5 seconds of a stranger's behavior can result in reasonably reliable inference of psychopathic personality, characterized by callousness, manipulation, and persistent antisocial behavior (Fowler, Lilienfeld, & Patrick, 2009). Observers may even be unaware of the specific nonverbal behavior that guides their evaluations of credibility. In the Canadian case *R. v. Lifchus* (1997), Justice Cory noted:

> It may be that the juror is unable to point to the precise aspect of the witness's demeanor which was found to be suspicious... A juror should not be made to feel that the overall, perhaps intangible, effect of a witness's demeanor cannot be taken into consideration in the assessment of credibility.

Second, expectancies about the truthfulness of a person may influence the observer's attention. For example, analyses of police interviews in England showed that the police interviewers were "certain" of the suspect's guilt before interviewing him or her in 73% of the cases (Moston, Stephenson, & Williamson, 1992). Saul M. Kassin (2005, p. 216), who had asked numerous American police officers whether they are

concerned that their persuasive interrogation methods may evoke false confessions, reported that the most common reply is "No, because I do not interrogate innocent people." When lying is expected, police officers may have little interest in listening to a suspect's flat denials and prefer to look at bodily signs to confirm deceit (Millar & Millar, 1998).

Third, formulating and asking the best questions in some contexts, particularly suspect interviews, can be a cognitively taxing task. Concurrent attempts to detect deceit during these interviews may further increase the cognitive demands on the interviewers (Patterson, 1995, 2006). Accordingly, interviewers may be inclined to detect deceit via nonverbal channels, because the processing of nonverbal cues requires fewer cognitive resources than the processing of verbal cues (Reinhard & Sporer, 2008).

Fourth, the preference for nonverbal behaviors as indicators of deception may result from training, which encourages such an emphasis. For example, police training manuals place greater emphasis on nonverbal cues than on speech-content cues as cues to deceit (for a review of visual cues mentioned in police manuals, see Vrij & Granhag, 2007). This nonverbal dominance is further emphasized with explicit statements. For example, Inbau et al. (2001) stated in their widely used training manual that "as much as 70 percent of a message communicated between persons occurs at the nonverbal level" (p. 143). Popular books by academics may also promote a reliance on nonverbal behaviors in catching liars. For example, in Paul Ekman's (1985/2001) book *Telling Lies: Clues to Deceit in the Marketplace, Politics and Marriage*, there is much greater attention to nonverbal cues of deception than to speech-related ones. Although this was probably justified when the first edition of the book was published in 1985, the past 25 years have witnessed the generation of a large body of speech-related deception research, particularly concerning criteria-based content analysis (for reviews, see Vrij, 2005, 2008a) and reality monitoring (for reviews, see Masip et al., 2005; Sporer, 2004; Vrij, 2008a).

This overemphasis on nonverbal cues to deception is problematic. Meta-analyses of verbal and nonverbal cues of deception have shown that many speech-related cues are more diagnostic of deception than are nonverbal cues (DePaulo et al., 2003; Vrij, 2008a). In addition, observers who pay sole attention to nonverbal cues are less accurate in discriminating truths and lies than are those who consider speech content (C.F. Bond & DePaulo, 2006; Burgoon, Blair, & Strom, 2008; Lindholm, 2008). In addition, paying attention to visual cues may encourage a lie bias, or tendency to judge someone to be a liar (C.F. Bond & DePaulo). An explanation for this is that people have stereotypical beliefs about the behavior of liars (e.g., gaze aversion, fidgeting) rather than of truth tellers (The Global Deception Team, 2006; Strömwall et al., 2004; Vrij et al., 2006). In other words, people can judge deception based on the presence of some cues, but they need to judge truthfulness based on the absence of some cues. People normally respond to the presence of a signal rather than to the absence of a signal. A lie bias heightens the risk of false suspicion, even conviction, of innocent suspects (Kassin, 2008a, 2008b; Kassin, Appleby, & Torkildson-Perillo, 2010; Kassin & Gudjonsson, 2004).

The Othello error

A common error in lie detection is to too readily interpret certain behaviors, particularly signs of nervousness, as diagnostic of deception. A common mistake for lie detectors is the failure to consider that truth tellers (e.g., an innocent suspect or defendant) can be as nervous as liars. Truth tellers can be nervous as a result of being accused of wrongdoing or as a result of fear of not being believed, because they too could face negative consequences if they are not believed (C.F. Bond & Fahey, 1987; Ofshe & Leo, 1997). The misinterpretation of signs of nervousness in truth tellers as signs of deceit is referred to as the *Othello error* by deception researchers (Ekman, 1985/2001), based on Shakespeare's character. Othello falsely accuses his wife Desdemona of infidelity, and he tells her to confess because he is going to kill her for her treachery. When Desdemona asks Othello to summon Cassio (her alleged lover) so that he can testify her innocence, Othello tells her that he has already murdered Cassio. Realizing that she cannot prove her innocence, Desdemona reacts with an emotional outburst, which Othello misinterprets as a sign of her infidelity. The Othello error is particularly problematic in attempting to identify high-stakes lies because of the observer's sense of urgency and a host of powerful cognitive biases that contribute to tunnel-vision decision making (see Porter & ten Brinke, 2009).

The use of heuristics

Instead of carefully scrutinizing someone's responses in evaluating his or her credibility, observers may rely on general decision rules (Fiedler & Walka, 1993). Person-perception researchers have observed that this can be an effective way for observers with limited time and attentional resources to deal with complex environments or demands (Albrechtsen, Meissner, & Susa, 2009; Macrae & Bodenhausen, 2001). However, general decision rules, or *heuristics*, can easily lead to systematic errors in decision making (Burgoon et al., 2008).

In the subsequent section, we review some heuristics that may lead to systematic errors when trying to detect deception. It should be noted, however, that there is a relatively recent wave of research that has challenged the view that relying on heuristics is necessarily bad. For example, since the mid-1990s, research has provided empirical support that the use of certain heuristics in certain contexts leads to effective, accurate decisions (Gigerenzer, Todd, & the ABC Research Group, 1999). Detecting deception can be a complex endeavor. Sometimes, observers have little time or information to formulate an informed decision, and they must rely on heuristics (consider, for example, a bank clerk confronted by a robber with one hand in his or her pocket and claiming to have a gun). The question then is *which* heuristics to use and which to avoid. Deception researchers have focused considerable attention on

problematic heuristics but little on potentially effective heuristics.

Several heuristics that are commonly used in assessing credibility can be identified. Because people encounter more truthful than deceptive messages in their daily lives, they assume that most behavior that they encounter is associated with honesty (i.e., the *availability heuristic*, O'Sullivan, Ekman, & Friesen, 1988), in stark contrast with the bias evidenced by police officers. A related heuristic is the *anchoring heuristic* (Elaad, 2003), referring to the tendency to make insufficient adjustments from an initial value or assessment (the anchor) resulting in a final decision that is biased toward this value. Thus, if observers are preoccupied in thinking that someone is telling the truth, they will make insufficient adjustments when contrasting evidence emerges. It has further been argued that as romantic relationships become more intimate, partners develop a strong tendency to judge the other as truthful, the so-called *relational truth-bias heuristic* (D.E. Anderson, Ansfield, & DePaulo, 1999; Stiff, Kim, & Ramesh, 1992). An opposite anchoring problem has been observed in the legal system. According to dangerous decisions theory (Porter, Gustaw, et al., 2010; Porter & ten Brinke, 2009), the reading of a suspect's or defendant's face and emotional expressions (the anchor) plays a powerful role in influencing decisions concerning his or her honesty. This theory predicts that the human brain makes instantaneous inferences about trustworthiness that influence various aspects of interpersonal evaluation, including those about credibility and culpability. For example, jurors make strong but often inaccurate intuitive judgments of a defendant's general trustworthiness quickly upon seeing his or her face for the first time, with this initial intuitive assessment having a substantial influence on the manner in which the credibility of ensuing information from and about the individual is interpreted (Bar, Neta, & Linz, 2006; Porter et al., 2008; Todorov, 2008).

The *probing heuristic* (Levine & McCornack, 1996a, 1996b, 2001) refers to observers' tendency to believe a source more after the source has been probed. Guided by the belief that probing is an effective lie-detection strategy, the source is more likely to be believed if probing does not result in clear signs of deceit (and it often will not). The *representativeness heuristic* (Stiff et al., 1989) refers to the tendency to evaluate a particular reaction as an example of a broader category. In the present context, it could explain people's inclination to interpret nervous behaviors as signs of deceit. The *consistency heuristic* refers to the tendency to judge consecutive consistent statements as being truthful and consecutive statements that are inconsistent as being deceptive (Granhag & Strömwall, 2000a, 2000b). The *expectancy violation heuristic* (Vrij, 2004) refers to the tendency to judge reactions that seem odd according to conversation norms and have a low base rate (e.g., keeping the eyes closed, or conversely, staring intently during a conversation) as being deceptive. According to the *falsifiability heuristic*, messages that that are easily falsifiable via reality checks appear less credible than messages that are not easily falsifiable, such as feelings, preferences, attitudes, and opinions (Fiedler & Walka, 1993).

The *facial appearance heuristic* (Vrij, 2004) refers to the tendency to judge people with attractive, symmetrical faces or baby-faced appearances as honest, and people with certain facial characteristics suggesting anger and unkindness as dishonest (Porter, England, Juodis, ten Brinke, & Wilson, 2008). Willis and Todorov (2006) found that observers infer the trustworthiness of others almost instantaneously upon seeing the face (100 milliseconds of exposure) and do so with a high level of confidence. Yet, Porter et al. (2008) found that observers were unable to discriminate philanthropists from felons featured in the television program *America's Most Wanted* despite believing that they "knew" who were the most and least trustworthy. Similarly, there are some faces that people agree look like that of a rapist, robber, or murderer (R. Bull & McAlpine, 1998; Dumas & Testé, 2006), which will influence the observer's assessment of honesty concerning the alleged offense.

The *visual cue primacy heuristic* (e.g., Burgoon et al., 2008; Stiff et al., 1989) refers to a tendency to assign primacy to visual information when attempting to detect deceit. Last, we add to this list the *single cue heuristic*, the oversimplified belief that all liars under all circumstances can be identified via single clear-cut cues. The belief that "liars look away" is probably the most popular example in this category (the *gaze aversion heuristic*; The Global Deception Team, 2006; Porter & ten Brinke, 2010).

Neglect of interpersonal differences

Obviously, there are large individual differences in people's behavior and speech (DePaulo & Friedman, 1998). Some people typically make many movements, others do not; some people are eloquent, others are not; some people show large variations in physiological responses, others do not, and so on. Although verbal lie-detection tools such as statement validity assessments attempt to control for these interpersonal behavioral differences via a validity checklist (Vrij, 2005, 2008a), assessing the impact of these individual differences remains a difficult task. Take, for example, controlling for *susceptibility to suggestion*, one of the factors appearing on the checklist. Some interviewees are more prone to an interviewers' suggestions than are others. The danger of suggestibility is that a suggestible person may be inclined to provide information that confirms the interviewer's expectations but that, in fact, is inaccurate. If the suggestible person is aware that the information that he or she provides is inaccurate, he or she is lying. Accordingly, Yuille (1988) and Landry and Brigham (1992) have recommended asking the interviewee a few misleading questions at the end of the interview to assess his or her susceptibility to suggestion. Because asking such questions about central information could harm the statement (it could contaminate someone's memory; Loftus, 2005; Loftus & Palmer, 1974; Porter, Yuille, & Lehman, 1999), Yuille (1988) recommends focusing on peripheral information (e.g., "When you were with your sister, which friend was also there, Claire or Sarah?" when the

interviewer is aware that there was no friend present). However, being restricted to asking questions about peripheral information is problematic because interviewees show more resistance to suggestibility for central aspects of an event than for peripheral aspects of an event (Dalton & Daneman, 2006; Goodman, Rudy, Bottoms, & Aman, 1990; Porter, Spencer, & Birt, 2003), and they are more resistant to suggestibility for stressful events, most likely the central information, than for less stressful events, most likely the peripheral information (Davies, 1991; Porter & Peace, 2007). Therefore, insight into interviewees' suggestibility for peripheral parts of the event cannot be effectively used to draw conclusions about their suggestibility for core events.

Nonetheless, professionals using statement validity assessments at least attempt to control for individual differences. Often, observers do not make such attempts when evaluating behavioral responses (Vrij, 2008a). Accordingly, people whose natural behavior looks "suspicious" (e.g., they are fidgety) run the risk of being falsely accused of lying. The literature provides examples of nervous-looking people whose nervousness led to being falsely accused. For example, in Florida, Tom Sawyer was interrogated for 16 hours regarding a sexual assault and murder and was issued threats, after which he gave a confession that likely was false. He became a prime suspect because he appeared embarrassed and his face flushed during an initial interview in which he denied involvement in the crime (Meissner & Kassin, 2002). In a notorious Canadian case, 14-year-old Steven Truscott was falsely convicted for the 1959 rape and murder of Lynn Harpur. In an initial interview with the suspect, inspector Graham observed that Truscott acted nervously and described him as a "lying, sexual deviant," initiating a process of tunnel vision that led to the boy's conviction and death sentence, later overturned (Porter & ten Brinke, 2009).

The tendency to interpret nervous behaviors as suspicious without taking individual differences into account puts several groups of people at risk, including introverted individuals and people who are socially anxious. The social clumsiness of introverts and the impression of tension, nervousness, or fear that is naturally given off by socially anxious individuals (DePaulo, Epstein, & LeMay, 1990; Riggio, Tucker, & Throckmorton, 1988; Schlenker & Leary, 1982) may be interpreted by observers as indicators of deception.

Errors are also easily made when people of different ethnic backgrounds or cultures interact, because behaviors naturally displayed by members of one ethnic group or culture may appear suspicious to members of another ethnic group or culture. Nonverbal behavior is culturally mediated. For example, Black Americans display more gaze aversion than do White Americans (Johnson, 2006a, 2006b; LaFrance & Mayo, 1976, 1978), and people from Turkey and Morocco who are living in the Netherlands show more gaze aversion than do native Dutch people (Van Rossum, 1998; Vrij, Dragt, & Koppelaar, 1992). It thus appears that looking into the eyes of the conversation partner is typical Caucasian behavior that is often not displayed by non-Caucasian individuals. Differences in culture contribute to this effect. Looking into the eyes of a conversation partner is regarded as polite in Western cultures but is considered to be rude in several other cultures such as, for example, Japan (Vrij & Winkel, 1991; Vrij, Winkel, & Koppelaar, 1991; Winkel & Vrij, 1990). Many groups of Aboriginals in Canada suppress expressions of their emotions, and such apparent flat affect may be considered inconsistent with the context at hand, and it may be interpreted as a sign of deception or lack of remorse by decision makers (Porter & ten Brinke, 2009). Brant (1993, p. 261) observed that most Caucasian Canadians see "people who do not provide direct eye contact ... as being shifty, devious, dishonest, crooks, slippery, untrustworthy, etc." In contrast, most Aboriginal cultures in Canada consider direct, sustained eye contact as rude, hostile, and intrusive. That is, the Aboriginal custom of avoiding eye contact as a sign of respect may easily be interpreted as an indication of deception by non-Aboriginal observers, including members of the judiciary.

Researchers have found other culturally determined differences in nonverbal behavior. For example, in the Netherlands, an experiment examining the nonverbal behavioral patterns of native Dutch Caucasian and Black Surinamese residents (citizens originated from Suriname, a former Dutch colony, but now living in the Netherlands) revealed large behavioral differences between the two groups, regardless of whether they were telling the truth or lying. Surinamese people made more speech disturbances, exhibited more gaze aversion, smiled more, and made more self-adaptors (e.g., fidgeting) and illustrators whether lying or not (Vrij & Winkel, 1991). In the United States, Johnson (2006a, 2006b) reviewed 120 videotaped police–citizen interactions of a noncriminal nature. The findings replicated those of Vrij and Winkel (1991) in that Blacks displayed more gaze aversion, smiling, and hand gestures than did Whites.

This means that observers need to be careful in cross-cultural interactions and should interpret the nonverbal behaviors displayed by senders of a different ethnic origin in light of cultural differences (Ruby & Brigham, 1997; Vrij, 2008a). Experimental research has demonstrated that this does not always happen and that cross-cultural nonverbal communication errors occur. That is, nonverbal behavioral patterns that are typical for an ethnic group are interpreted by Caucasian observers as signs of deception (Vrij & Winkel, 1992, 1994). It is important to note that these issues are relevant not only for police investigators, but also for professionals working in the immigration service (Granhag, Strömwall, & Hartwig, 2005).

Neglect of intrapersonal variations

Different people respond differently not only in the same situation (*inter*personal differences), but also in different contexts (*intra*personal differences). Neglecting or underestimating intrapersonal differences is another error that lie catchers make. In police interviews, detectives are advised to examine a suspect's natural, truthful behavior during the small talk preceding the interview and to compare this behavior with the behavior

shown by the suspect during the actual interview. Differences in behavior could then be interpreted as "significant" (Inbau et al., 2001). This approach is also used and advocated by researchers (Frank, Yarbrough, & Ekman, 2006; Hirsch & Wolf, 2001). Although the approach sounds appealing, it is conducive to forming incorrect judgments because it is based on an incongruent comparison. Engaging in small talk and discussing the crime itself are fundamentally different situations. Small-talk conversations are low-stakes situations in which the suspect's responses are unlikely to have any negative consequences. In contrast, the core investigative elements of the interview are high-stakes situations in which the suspect's reactions and responses are critical. Therefore, it is not surprising that both guilty and innocent suspects tend to show different behaviors during small talk compared to during the actual interview (Vrij, 1995). This problematic issue also plagues the control-question polygraph test, because it is difficult to come up with control questions that are as significant as the key questions concerning the crime (National Research Council, 2003). The tendency to neglect or underestimate the importance of intrapersonal differences is an error that not only lie detectors make; it is a well-known error in social perception and relates to the fundamental attribution error (Ross, 1977).

Existing interview techniques

Many interview strategies advocated by police manuals can impair lie detection. For example, police detectives are sometimes advised to confront suspects at the beginning of the interview with the evidence they have previously in their investigation (Hartwig et al., 2006; Leo, 1996). This tactic is designed to show suspects that it is fruitless to remain silent and that they are better off confessing. Experimental research has revealed that this interview style hampers lie detection (Hartwig, Granhag, Strömwall, & Vrij, 2005). One of the problems liars can face is ignorance about the level of knowledge held by the observer. This makes it difficult to know what they can say without assuming the risk of offering statements that are contradictory with known facts. If police officers promptly disclose their knowledge, they reduce the uncertainty for deceptive suspects and may inadvertently facilitate the ease of lying. Disclosing evidence early on provides liars with the opportunity to change their stories and to give an innocent explanation for the evidence.

Another misguided strategy from an informed lie-detection perspective is to accuse someone of lying. This affords deceptive suspects the ideal opportunity to "escape" from the interview situation by saying that they will no longer cooperate with the investigation, claiming that further cooperation is futile because they are not believed anyway. Also, accusing someone of lying may elicit the same responses in liars and truth tellers. That is, both suspects correctly accused of lying and those wrongly accused of lying may become afraid of not being believed (Ofshe & Leo, 1997). Because of that fear, both groups may show the same nervous responses (C.F. Bond & Fahey, 1987).

Overconfidence in lie-detection skills

The final error that we will highlight is that professional lie catchers tend to overestimate their ability to detect deceit. Research has consistently shown that when professional lie catchers and laypersons are compared, professionals are more confident in their veracity judgments but are no more accurate (DePaulo & Pfeifer, 1986; Garrido, Masip, & Herrero, 2004; Kassin, Meissner, & Norwick, 2005; Meissner & Kassin, 2002). This tendency to overconfidence is not unique to police officers but is common among many groups of professionals in carrying out their job duties (Allwood & Granhag, 1999). Further, some research has suggested that more experienced professional lie catchers are more confident in their credibility-assessment abilities than are their less experienced counterparts but that they are no more accurate (e.g., Porter et al., 2000).

The overconfidence could, in part, be explained by overzealous promotion of lie-detection tools by those with commercial interests. No lie-detection tool used to date that is based on analyzing nonverbal and verbal behavior is accurate—far from it (Vrij, 2008a). Despite the fallibility of those tests, Paul Ekman, an American emeritus professor of psychology who has specialized in nonverbal cues to deceit, said in an interview with *The New York Times* (Henig, 2006) that his system of lie detection can be taught to anyone, with an accuracy of more than 95%. However, there is no published study that supports this claim. In a similar vein, one of the interview techniques discussed in detail in Inbau et al.'s (2001) manual is the behavior analysis interview. The authors claimed that interviewers specifically trained and experienced in behavior analysis assessment can correctly identify the truthfulness of a person 85% of the time. However, conclusive evidence to support this claim is lacking (Blair & Kooi, 2004; Horvath, Jayne, & Buckley, 1994; Vrij, Mann, & Fisher, 2006a; Vrij, Mann, Kristen, & Fisher, 2007).

Confidence in lie detection is not related to accuracy. In a meta-analysis of the confidence–accuracy relation that included 18 samples, the relation appeared to be virtually nonexistent ($r = .04$), not differing significantly from zero (DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997). Such a low correlation between confidence and accuracy is not unique for veracity judgments; other areas of cognitive performance, such as eyewitness identification, reveal a similar pattern (Sporer, Penrod, Read, & Cutler, 1995).

High confidence in one's ability to catch liars can be harmful when the confidence is unjustified (Kalbfleisch, 1992). High confidence often results in making quick decisions on the basis of limited information (Levine & McCornack, 1992; Lord, Ross, & Lepper, 1979), or tunnel vision (Porter & ten Brinke, 2010). In addition, high confidence may make investigators attempt to detect lies via demeanor alone and not search for physical evidence (Colwell, Miller, Lyons, & Miller, 2006). High confidence also is likely to reduce motivation to learn more about lie detection, because investigators may consider themselves already experts in the area. An unwillingness to learn more about lie detection is obviously undesirable, given professional lie catchers' typically low performance at the task

(C.F. Bond & DePaulo, 2006; Vrij, 2008a). Regarding this performance, Vrij reviewed 28 lie-detection studies with professionals (e.g., police officers, police detectives, parole officers) as lie detectors. On average, these professionals correctly classified 56% of liars and 56% of truth tellers, whereas 50% could be expected by chance alone. A lively discussion about the existence of individual differences in the ability to detect deceit has recently emerged.[4]

Overconfidence is a problem not only when it comes to one's general ability to detect lies but also when it leads to serious problems in an individual veracity assessment. For example, overconfidence in assessing a denying (but guilty) suspect as a truth teller will result in the suspect being released, and it provides opportunities for the suspect to commit more crimes. In addition, if a police detective is confident that a suspect is lying, he or she may subject the suspect to persuasive interrogation techniques in order to obtain a confession. This can harm innocent suspects in particular. Kassin, Goldstein, and Savitsky (2003) found that when innocent suspects are mistakenly identified as guilty, an interrogation style that is even more coercive than those experienced by guilty suspects can occur. That is, interrogators who do not believe the innocent suspect's denials are inclined to double their efforts to elicit a confession (Kassin et al.).

Opportunities in Lie Detection

Avoiding the errors

Avoid examining the wrong cues and pay attention to the more diagnostic verbal and nonverbal cues to deceit. As previously discussed, observers often base their veracity decisions on cues that are not diagnostic of deception. Thus, it sounds plausible that observers may become better at discriminating truths and lies if they are taught to pay attention instead to deception cues that are more diagnostic. Several training studies have addressed this issue, and these are reviewed in detail by Frank and Feeley (2003) and Vrij (2008a).

In all extant training studies, observers have been exposed to short videotaped or audiotaped interviews with a number of people who were telling either truths or lies. Generally, 1 of 3 procedures was used. Some studies have used a *focusing* procedure in which observers are asked to pay attention to specific cues and ignore others. Other studies have used an *information* procedure in which observers receive information about the actual relation between certain behaviors and deception. Yet other studies have used an *outcome feedback* procedure in which each time observers made a decision, they are informed about the accuracy of that decision. In all three types of procedures, the performance of these trained participants is then compared with the performance of untrained and uninformed (control) participants.

Most studies have revealed that trained observers are better at distinguishing between truths and lies than are control observers, regardless of the training method used. However, these improvements have typically been small. On average, the control observers detected 53.4% of the truths and lies correctly, and the trained observers 57.66%. In other words, people can, to a limited extent, be trained to become better lie detectors.

The training studies have revealed two more outcomes that are worth discussing. First, Levine, Feeley, McCornack, Hughes, and Harms's (2005) experiment included bogus training groups that were taught cues that are *not* diagnostic cues to deception. They found that sometimes these bogus training groups performed better than the control groups, suggesting that the simple act of training, rather than the content of the training, may improve accuracy. In alignment with this, Porter, Woodworth, McCabe, and Peace (2007) found that the provision of *any* feedback (accurate or inaccurate) following deception judgments had a positive, albeit modest, influence on deception detection. It could be that the trained observers assessed the messages more critically than the control observers (Levine et al., 2005). Alternatively, training may make observers more motivated to perform well (Hartwig & Bond, 2010).

Other studies have showed worse performance by trained observers than by control observers. For example, when Kassin and Fong (1999) trained observers to examine the cues taught by the Inbau group as reported in their manual (Inbau et al., 2001), the observers performed worse than their untrained counterparts. In other studies where it was found that training impaired lie detection (Köhnken, 1987; Vrij, 1994; Vrij & Graham, 1997), the observers were police officers rather than undergraduate students. Vrij and Graham found that the students performed better as a result of the information they received, whereas police officers performed worse after having received the same information. We can only speculate as to why police officers do not appear to benefit from the provision of such information. One explanation is that the information confuses them (see also Köhnken). Perhaps the information Vrij and Graham gave about the relation between personality traits and deceptive behavior was beyond the grasp of the police officers who are probably not familiar with personality theories. The student observers in their experiment were psychology students and hence familiar with personality theories (albeit not with the relation between personality traits and deception). A second explanation is that police officers refused to use the information provided because it contradicted their own beliefs. For example, in Vrij's (1994) study, the observers were told that liars typically show a decrease in hand and finger movements, whereas police officers typically assume that an increase in hand and finger movements indicates deception. Perhaps the officers refused to accept the information provided by an outsider (the experimenter) and continued to rely on their own experience and beliefs instead.

The small improvements found in research may not necessarily reflect the true potential of teaching people to detect deceit. The training programs were typically brief and sometimes lasted no more than 15 minutes. Longer, more intensive training sessions such as the ones used in Porter et al.'s (2000) study (2-day training: pretraining vs. posttraining, 40.4% vs. 76.7%)

and in Porter, Juodis, ten Brinke, Klein, and Wilson's (2010) study (2-hour training: pretraining vs. posttraining, 51.2% vs. 60.7%) achieved greater success. The training programs also did not address the complex nature of lie detection. For example, in studies using the information procedure, observers were taught a set of cues that liars may display. This approach is limited because not all liars will show these specific sets of cues. Moreover, in all of these studies, the observers were exposed to low-stakes truths and lies, and low-stakes situations do not provide much opportunity to detect deception. It could thus be possible that training has larger effects if observers are given more sophisticated training and are exposed to truths and lies told in high-stakes situations.

We believe, however, that training programs as described in this section will never yield high accuracy rates. The limitation of these programs is that trainees are restricted to passive observation of truth tellers and liars. Such a method is limited because cues of deception are faint and unreliable. We therefore see more potential in training programs that teach trainees to actively *elicit* or *enhance* diagnostic cues to deception. In the section on "Exploiting the Different Mental Processes of Truth Tellers and Liars," we present interview styles designed to achieve this.

Avoid relying on nonverbal cues only. Research addressing the individual strategies of lie detectors has indicated that detecting truths and lies becomes more successful when speech content is taken into account. Mann et al. (2004) showed 99 police officers 54 videotaped fragments of police interviews with murderers, rapists, and arsonists and found that good lie detectors reported to have relied upon verbal cues (e.g., vague reply, contradictions in story) more often than did poor lie detectors. In addition, there was an inverse relation between the number of visual cues reported to have been relied upon (e.g., gaze aversion, posture, movements) and accuracy. In particular, police officers who mentioned that liars look away and fidget achieved the poorest scores. In other words, those who listened carefully to what suspects had to say were better lie detectors than those who concentrated on suspects' nonverbal behavior.

D.B. Anderson, DePaulo, Ansfield, Tickle, and Green (1999) and Feeley and Young (2000) found a positive relation between the number of vocal cues that participants reported to have relied upon (e.g., speech errors, speech fillers, pauses, voice) and accuracy. In a study in which participants attempted to detect truths and lies told by a convicted murderer, participants who mentioned gaze aversion and fidgeting as cues to deceit achieved the lowest accuracy scores (Vrij & Mann, 2001a). Also, Porter et al. (2007) found that the more visual cues the participants reported, the worse their ability to distinguish truths and lies. In summary, all of these studies showed that in order to detect lies, listening carefully to what is said is necessary and that merely paying attention to behavior impairs lie detection.

Another body of research suggests that a "holistic" approach to detecting deception may be ideal. Ekman and O'Sullivan (1991) found that participants who reported to have relied upon both vocal/verbal and visual cues obtained higher accuracy rates than did participants who reported to have relied upon only vocal/verbal or visual cues. This is supported by experimental research in which the nonverbal and verbal cues of truth tellers and liars were examined. That research has demonstrated that the best classifications of truths and lies are made when both sets of cues are taken into account (Porter & Yuille, 1996; Porter et al., 1999; Vrij, Akehurst, Soukara, & Bull, 2004a; Vrij, Edward, Roberts, & Bull, 2000; Vrij, Evans, Akehurst, & Mann, 2004). Thus, attendance to multiple cues from words and the visual channel should provide the lie catcher with better ammunition for the task at hand (Porter & ten Brinke, 2010).

Observers can pay attention to nonverbal behavior and speech simultaneously in three different ways, which all enhance lie detection. First, observers could take into account both nonverbal and verbal cues without looking at the relation between the two sets of cues. This was the case in the previously discussed research. Second, observers could examine nonverbal behavior in relation to speech content, an approach common in communication research (Bavelas & Chovil, 2006; Bavelas, Chovil, Coates, & Roe, 1995; Bavelas & Gerwing, 2007; P. Bull, 2009; Freedman, 1972; Kendon, 1994, 2004; McNeill, 1985, 1992) but often ignored by deception researchers. A recent experiment showed the potential of this approach (Caso, Maricchiolo, Bonaiuto, Vrij, & Mann, 2006). When the entire interview was taken into account, truth tellers and liars displayed a similar number of illustrators. Differences did emerge between truth tellers and liars only when specific types of illustrators were examined when answering specific questions. Third, observers could examine mismatches between nonverbal behavior and speech content (Ekman, 1985/2001; Ekman & O'Sullivan, 2006). Thus, a person who makes a head shake while agreeing to cooperate may not actually be as cooperative as he or she wants to appear. Thus, although a perfectly reliable cue to deception does not exist, the combination of attention to changes in nonverbal/body language, verbal, and facial channels—ideally videotaped to permit review and systematic analysis—can provide the basis for an informed opinion about credibility as long as it is backed by other evidence (Porter & ten Brinke, 2010).

However, mistakes are easily made. For example, some people display clear signs of distress when they talk about a negative event they have experienced, whereas others do not (Burgess, 1985; Burgess & Homstrom, 1974; Vrij & Fischer, 1995). Thus, the varying communication styles represent a personality factor (Littman & Szewczyk, 1983). However, observers, including police detectives, typically believe that absence of distress during an interview about an upsetting event is a valid indicator of deceit (Greuel, 1992). As a result, different emotional displays have a differential effect on the perceived credibility of complainants, and emotional victims are more readily believed than victims who report their experience in a controlled manner (Baldry & Winkel, 1998; Baldry, Winkel, & Enthoven, 1997; Bollingmo, Wessel, Sandvold, Eilertsen,

& Magnussen, 2009; Bothwell & Jalil, 1992; Kaufmann, Drevland, Wessel, Overskeid, & Magnussen, 2003; Vrij & Fischer, 1997; Wessel, Drevland, Eilertsen, & Magnussen, 2006; Winkel & Koppelaar, 1991).

Another relevant point relating to the potential for misinterpretation by the lie catcher concerns facial expressions. Ekman has long argued that deceptive emotional information is betrayed (leaked) by *microexpressions*, fleeting but complete facial expressions that are thought to reveal the felt emotion during emotional concealment and are suppressed within 1/5th to 1/25th of a second (Ekman, 1985/2001). This idea has enjoyed increasing popularity in the media (Henig, 2006) and scientific community (Schubert, 2006), despite being backed by little empirical research. Porter and ten Brinke (2008) conducted the first thorough investigation of facial expressions associated with genuine and deceptive emotions. Participants viewed disgusting, sad, frightening, happy, and neutral images, responding to each with a genuine emotion or a deceptive one, by either *masking*, replacing one emotion with another, or *simulating*, creating an emotional expression in a neutral state, while being judged by "blind" observers. The researchers analyzed each 1/30-second frame (104,550 frames in 697 expressions) for the presence of the muscle actions of the universal expressions and for the presence of microexpressions. Their findings indicated that emotional expressions inconsistent with the intended display did occur more frequently in the masked condition than in the genuine or simulated conditions. All participants showed such predicted "leakage" on at least one attempt at faking an emotion. However, Porter and ten Brinke found only a small number of partial (lower or upper face) microexpressions. Although some of the microexpressions betrayed the hidden emotion, they sometimes occurred during genuine expressions. The leakage was typically longer and more salient than Ekman had predicted. As such, the lie catcher should attend to the expressions that are inconsistent with what is being said or with the context, but he or she should be cognizant that these expressions can be meaningless. Therefore, when lie detectors believe that there is a mismatch between someone's nonverbal behavior and speech content, they should be careful about how to interpret it. A final judgment that the person is lying should not be made too quickly, and alternative explanations should be considered. In this context, some researchers refer to these cues as "hotspots" deserving further attention rather than as being necessarily indicative of lying (Frank et al., 2006, p. 234). There is a serious risk that nonverbal hotspots are too easily interpreted as lies. In that context, we underline Porter and ten Brinke's (2010) conclusion that nonverbal cues only assist investigators who are informed about the complex relations between behavior and deceit.

Avoid the Othello error: Consider alternative explanations when interpreting cues of emotions and cognitive load. As previously mentioned, the *Othello error* refers to mistakenly interpreting signs of nervousness as cues to deceit. The difficulty that lie detectors face is that both liars and truth tellers may display signs of emotions and/or nervousness in high-stakes situations. Consider the distress one must feel to be falsely accused by the police of having committed a serious crime or by a partner about having had an affair. Emotion cues may not conclusively demonstrate that someone is lying, and the lie detector should thus be cautious in interpreting such cues as signs of deceit. Instead, in interpreting emotional responses, the lie detector should consider questions such as the following: "Is my questioning likely to evoke emotions in the respondent, regardless of whether he or she is guilty?" "Is the present situation likely to evoke emotions in the respondent anyway?" And "Is this person the type who is likely to be emotional in this situation anyway?" (Ekman, 1985/2001).

In theory, another cluster of cues could betray deception—cues associated with having to think hard (labeled *cognitive load*; Buller & Burgoon, 1996; DePaulo et al., 2003; Ekman, 1985/2001; Vrij, 2008a). For example, Porter and ten Brinke (2008) found that when participants worked hard to neutralize an emotion (e.g., maintaining a neutral expression when viewing a horrific accident scene), their blink rate lowered relative to when they expressed a genuine emotion (e.g., showing fear or horror when viewing the same scene). A decrease in blink rate is a sign of cognitive load (Bageley & Manelis, 1979). In forensic settings, however, such cues are not solely exhibited by liars; truth tellers may have to think hard while answering questions in a cognitively and emotionally complex context. Again, in interpreting cues of cognitive load, the lie detector should ask him- or herself the same kinds of questions as when interpreting signs of emotions, such as "Is my questioning likely to evoke cognitive load in the respondent, regardless of whether he or she is guilty?"

Avoid relying on heuristics and rely on multiple cues in a flexible manner. As previously discussed, deception research has revealed that no single behavioral or verbal cue is uniquely related to deception. In other words, there is no giveaway clue like Pinocchio's nose. Instead, different people show different cues to deception in a given situation (i.e., interpersonal differences) and the same person shows different cues to deception on different occasions (i.e., intrapersonal differences). Therefore, it is inappropriate to use fixed decision rules on the basis of heuristics such as "liars look away" when attempting to detect deceit. In fact, research has demonstrated that people who focus on single nonverbal or verbal cues are typically poor lie detectors (Mann et al., 2004; Vrij & Mann, 2001a). Instead, it is better to make veracity assessments on the basis of multiple cues (Ekman, O'Sullivan, Friesen, & Scherer, 1991; Porter & ten Brinke, 2010; Vrij et al., 2004a; Vrij, Edward, et al., 2000; Vrij & Mann, 2004). However, even such clusters of cues do not fit all liars; they also do not fit a particular liar in all situations. In other words, fixed decision rules that include multiple cues are not satisfactory either. Instead, better accuracy rates are achieved by using flexible decision rules that include multiple cues (Ekman & O'Sullivan, 1991; Ekman, O'Sullivan, & Frank, 1999; Mann et al., 2004; Vrij, 2008a).

Take into account inter- and intrapersonal differences and pay attention to deviations from a person's honest reactions in similar situations: The comparable truth. Lie detectors should take inter- and intrapersonal differences into account when making veracity judgments. Therefore, the relevant question for the lie detector to ask is whether the nonverbal behavior and speech patterns displayed by a person differ from this person's known behavior when delivering truthful responses. As discussed earlier, we advise police detectives to examine a suspect's natural truthful behavior during the small-talk preceding the interview and to compare this behavior with the behavior displayed by the suspect during the actual interview. This approach is prone to incorrect judgments, because engaging in small talk and discussing the crime are two fundamentally different situations. For this technique to work, it is essential that the known truthful response (e.g., baseline response) is made under similar conditions to the response under investigation, labeled the comparable truth (Vrij, 2008a). People react differently in formal settings (e.g., during a selection interview) than in informal settings (e.g., when at home with the family). According to Vrij (2006), they also react differently when they are accused of wrongdoing (e.g., situation during the actual interview) than when they are unchallenged (e.g., situation during small talk), and they respond differently in high-stakes situations than in low-stakes situations (Porter & ten Brinke, 2010; Vrij, 1995). In addition, people show different behaviors when they are interviewed by different people (Vrij & Winkel, 1991). Behavior is also topic related: People respond differently when discussing a topic that embarrasses them than they do when discussing a neutral topic (Kleinke, 1986), and they respond differently when discussing a topic that they care about or is important to them than they do when discussing a topic with which they have less personal involvement (Davis & Hadiks, 1995; Matarazzo, Wiens, Jackson, & Manaugh, 1970). Last, people's behavior sometimes varies over time in the same interview (Buller & Burgoon, 1996; Burgoon et al., 1999; Stiff, Corman, Krizek, & Snider, 1994; White & Burgoon, 2001), or, if they are interviewed on more than one occasion, changes may occur over repeated interviews (Granhag & Strömwall, 2002). Therefore, when lie detectors wish to compare a person's given nonverbal response with his or her truthful nonverbal response, they need to make sure that the given and truthful responses are taken from the same interview setting, that the person talks about similar topics in the given and truthful parts, and that these topics are discussed within a short period of time.

Vrij and Mann (2001a) provided an example of how the comparable-truth technique could be used. During a videotaped real-life police interview, a man suspected and later convicted of murder was asked to describe his activities during a particular day. The murder suspect described his activities during the morning (went to work), afternoon (visited a market) and evening (visited a neighbor). Detailed analyses of the videotape revealed a sudden change in behavior as soon as he began to describe his activities during the afternoon and evening. A possible reason for this variation may have been that he was lying, a view supported by the evidence. Police investigators could confirm his story with regard to his morning activities, but they revealed that his statement about the afternoon and evening was fabricated. In reality, he met the victim in the afternoon and killed her later that day. In this case, we were able to make a good comparison. The man described a seemingly normal day, and there are no good reasons why different behaviors would emerge while describing different parts of that day.

The comparable-truth technique has inevitable shortcomings, and mistakes will still be made with its application. The main problem is that it is difficult to rule out that the observed nonverbal and verbal differences are caused by factors other than deceit. Open-mindedness when interpreting the differences in behavior and speech is thus crucial. Also, differences between the baseline behavior and speech and the behavior and speech under investigation may be subtle and therefore difficult to spot. Last, an absence of behavioral and speech-related differences between the baseline behavior and speech and those under investigation does not necessarily mean that the person is telling the truth.

Exploiting the different mental processes of truth tellers and liars

The first five guidelines share one feature: They all aim to examine and interpret more carefully the nonverbal and verbal cues displayed by liars. And they have one serious limitation: The cues that lie detectors are encouraged to examine and interpret are faint and unreliable. In this section we discuss a fundamentally different approach to nonverbal and verbal lie detection: to elicit more, more blatant, and more reliable cues to deceit. We achieve this aim by exploiting the different psychological states of truth tellers and liars via two different approaches. The first approach, *strategic questioning*, uses specific questions that elicit the most differential responses between truth tellers and liars. The second, *imposing cognitive load*, makes the interview setting more difficult for interviewees. We argue that this affects liars more than truth tellers, thereby resulting in more and more blatant differences between the two. Both approaches require interviewees to talk. Interviewees can be encouraged to talk via an information-gathering interview style, as discussed in the subsequent section.[5]

Use an information-gathering interview style. The police commonly use two types of interview styles: information-gathering and accusatory (Moston & Engelberg, 1993). In the information-gathering style, interviewers ask suspects to give detailed statements about their activities through open questions (e.g., "What did you do yesterday between 3 p.m. and 4 p.m.?" "You just mentioned that you went to the gym; who else was there?"). By comparison, in the accusatory style, interviewers confront suspects with accusations (e.g., "Your reactions make me think that you are hiding something from me."). Information-gathering interviews encourage suspects to talk, whereas accusatory interviews often yield short denials

(e.g., "I am not hiding anything"). Therefore, information-gathering interviews typically elicit more information about an event and result in longer responses than do accusatory interviews (Fisher, Brennan, & McCauley, 2002; Vrij, Mann, & Fisher, 2006b; Vrij et al., 2007).

An information-gathering interview style is desirable for lie-detection purposes for several reasons. A good lie-detection strategy is to check the factual information provided by an alleged liar with the available evidence. The provision of a high quantity of details, most likely to result from an information-gathering interview, permits more opportunities for the lie detector to identify inconsistencies and contradictions between the answer and available evidence. Second, information-gathering interviews result in more nonverbal cues to deceit than do accusatory interviews (Vrij, 2006), because longer stories afford more opportunities for nonverbal cues to deception to be displayed (DePaulo et al., 2003). In addition, being accused of wrongdoing (i.e., accusatory interview style) is likely to affect the behavior of both truth tellers and liars in a similar way, and the accusation can have a stronger effect on someone's nonverbal behavior than the act of lying itself (C.F. Bond & Fahey, 1987; Ofshe & Leo, 1997). Consequently, differences in nonverbal behavior between truth tellers and liars are overshadowed by the effects of the accusation.

The third advantage of conducting an information-gathering interview is that it also results in more verbal cues to deceit (Vrij et al., 2007). Longer stories afford more opportunities for verbal cues of deceit to occur, because words are the carriers of such cues. A criteria-based content analysis, for example, requires the availability of a story and is not possible with an outright denial. Fourth, information-gathering interviewing does not involve accusing suspects of any wrongdoing or other tactics designed to cause distress. It could be a safeguard against false confessions that can occur with coercive interview styles aimed at creating duress/distress (Gudjonsson, 2003; Kassin, Appleby, & Torkildson-Perillo, 2010). Fifth, veracity judgments in accusatory interviews are made with more confidence than are those in information-gathering interviews (Vrij et al., 2007), potentially leading to tunnel vision. If lie detectors monitor their confidence and do not become overzealous (which is known to impair lie-detection accuracy; Porter et al., 2007), they are more likely to defer making such conclusive judgments and gather more evidence (see also Levine & McCornack, 1992).

Although the information-gathering interview is a good start in discriminating truth and deceit, that approach alone is not sufficient to elicit diagnostic cues to deception (Granhag & Vrij, 2010; Vrij & Granhag, 2007). More sophisticated strategies incorporated within the information-gathering interview are needed and are discussed in the remaining part of this review.

The strategic-questioning approach: Ask unanticipated questions. A consistent finding in deception literature is that, when possible, liars prepare themselves for anticipated interviews (Granhag, Andersson, Strömwall, & Hartwig, 2004; Granhag, Strömwall, & Jonsson, 2003; Hartwig et al., 2007; Vrij et al., 2009). The act of planning and rehearsing a story can lead to vulnerabilities that investigators can consider. Rehearsal leads to overly scripted responses. One of the criteria of criteria-based content analysis with the greatest support in assessing credibility is unstructured reproduction (supported in at least 50% of relevant studies; see Vrij, 2008a). Truthful accounts tend to be more unstructured and less chronological than rehearsed deceptive accounts, which tend to be overly scripted and chronological (e.g., "I did this ... then this happened ... then I did this," and so on). A liar wants to keep his or her story straight (impression management) and will memorize the details of the story in order (Porter & ten Brinke, 2010).

Further, the effectiveness of a liar's planning strategy is limited, because it can only work when liars correctly anticipate the questions that will be asked. Investigators can exploit this limitation by asking questions that liars do not anticipate (e.g., spatial questions) or by asking questions in a format that liars do not anticipate (e.g., drawings).

In an empirical test of the unanticipated-questions technique, liars and truth tellers were interviewed individually about having lunch together at a restaurant (Vrij et al., 2009). Although the pairs of truth tellers did not have lunch together, the liars were instructed to pretend that they had. All pairs were given the opportunity to prepare for the interview. The interviewer asked typical opening questions that the interviewees later said they had anticipated (e.g., "What did you do in the restaurant?"), followed by questions about spatial details (e.g., "In relation to the front door and where you sat, where were the closest diners?") and temporal details (e.g., "Who finished their food first, you or your friend?") that the interviewees said they had not anticipated. Further, they were asked to draw the layout of the restaurant (unanticipated). On the basis of the overlap in responses to the anticipated opening questions between the individuals, the liars and truth tellers could not be classified at a level above chance. However, on the basis of the responses in the unanticipated questions, up to 80% of pairs of liars and truth tellers could be correctly classified, particularly when assessing drawings (i.e., the drawings were less alike for the pairs of liars than they were for the truth tellers). In summary, asking unanticipated questions about central topics leads to identifiable betrayals among liars.

Asking unanticipated questions can also be effective when assessing individual interviewees rather than pairs of interviewees. An interviewer could ask the same question twice in the same or different interviews. When liars have not anticipated the question, they have to fabricate an answer on the spot. A liar's memory of this fabricated answer may be more unstable than a truth teller's memory of the actual event. Therefore, liars may contradict themselves more than truth tellers may (Fisher, Vrij, & Leins, in press). This approach probably works best if the questions require detailed answers given in different formats. Truth tellers will have encoded the topic of investigation along more dimensions than will liars. As a result, compared with liars, truth tellers should be able to recall the event more flexibly (along more dimensions). Thus, the question "How old

are you?" followed by the question "What is your date of birth?" is more difficult to answer for liars than for truth tellers and results in longer latency periods in liars (Walczyk et al., 2005). In addition, when asked to verbally describe and sketch the layout of a restaurant, truth tellers' verbal answers and drawings show more overlap than do those of liars (Leins, Fisher, Vrij, Leal, & Mann, in press).

Another experiment showed further promise for the use of drawings as a lie-detection tool (Vrij, Leal, et al., 2010). The researchers sent 31 participants on a mission that included picking up a decoder from one agent and delivering it to a second agent. After delivering the decoder to the second agent, the participants were asked to (a) verbally describe what they had seen at the location where they had received the decoder and (b) sketch what they had seen at that location. Half of the participants were told to answer with a lie and half were told to answer with the truth. The liars were requested to pretend to have been on a different mission in which they received the decoder at a different location from a different agent. The results indicated that the drawings were more useful for lie detection than were the verbal accounts. Only 2 of 16 liars (12.5%) included the pretend agent from whom they claimed to have received the decoder in their drawing, whereas 12 of 15 truth tellers (80%) sketched the real agent from whom they had received the decoder. In their verbal descriptions, again 2 of 16 liars (12.5%) mentioned the pretend agent from whom they claimed to have received the decoder, whereas 8 of 15 truth tellers (53%) did mention the real agent. There are two possible reasons why liars were inclined to omit the pretend agent from the sketch and verbal description. First, since there was no actual agent present at the location they claimed to have received the decoder, they forgot to add an agent to their drawings and descriptions. Second, liars may be reluctant to include people in their drawings or verbal descriptions because it might trigger further questions about who those people actually were.

Why did more truth tellers sketch the agent (80%) than verbally described the agent (53%)? It may be hypothesized that after sketching the stable elements, the truth tellers may have noticed that the agent was missing from the drawing. After narrating the stable elements of the location, however, truth tellers will have been less aware of this omission because of difficulties in building a mental picture of a location on the basis of narratives. Future research could examine this hypothesis.

In a related vein, Liu et al. (2010) asked half of a group of children (10–12 years of age) to tell the truth about a self-experienced event and the other half to lie about such an event. The researchers found that lying children were more willing to answer odd questions (e.g., "Can you remember what you had in your left pocket when being stung by the bee?") than were truth-telling children, whereas no difference was found in the willingness to answer standard questions. Hence, asking unanticipated questions elicited a cue to deception (i.e., increased willingness to answer the impossible questions). The finding can be explained by acknowledging that the lying children had to *act* to appear honest, whereas truth-telling children did not have to do this. Liu et al. speculated that liars were afraid that an "I don't know" answer would sound suspicious. Hence, merely acting in an honest manner might result in some actions that are more rarely seen among those who are truly honest.

The strategic questioning approach: Ask temporal questions when suspecting a scripted answer. A good strategy for liars is to provide a story that is, in fact, true, but that happened at a different time than the time of interest (see the earlier section on embedded lies). For example, a guilty male suspect who denies involvement in a crime could claim that he was at the gym when the crime took place. If he is indeed familiar with the gym, he can now truthfully recall an experience there, describe its layout, the equipment that he uses there, and so on. The only fabricated part in this story is the time he was there. Lie detectors should be aware of this lying strategy. Questions about the layout of the gym and activities occurring are not necessarily effective because they enable liars to relate true experiences. Instead, questions should be asked that are specifically related to the particular time that the interviewee claims to have been where they say they were. For example, the interviewer could ask time-related questions about key events, such as which instructor was working at the time he or she claims to have visited the gym, who else was present, and so forth.

The specific question approach: The devil's advocate approach. Verbal lie-detection tools (such as statement validity assessments) are designed to distinguish between truths and lies when people describe events that they claim to have experienced. As a result, many assessment criteria focus on perceptual detail to examine what people report having seen, heard, felt, or smelled during these events. However, people lie not only about their experiences but also about their opinions. Determining the veracity of such conceptual representations may not be important in typical police suspect interviews because these are mainly concerned with detecting lies about transgressions. However, it can be important in many security settings such as, for example, when deciding whether an informant is (a) indeed as much anti-Taliban or against Muslim fundamentalism as he or she claims or (b) truly entering the United Kingdom or the United States solely for the purpose of university study. Incorrect veracity judgments can do irreparable harm in such situations, as demonstrated by the loss of seven CIA agents in Afghanistan on December 30, 2009. The CIA agents were killed via a suicide attack by a man they thought was going to give them information about Taliban and Al-Qaeda targets in Pakistan's tribal areas. The CIA agents had used polygraph tests to check the man's sincerity and were aware that he had posted extreme anti-American views on the Internet. However, it was decided that the views he had expressed were part of a good cover, and the possibility that they were his real views was discounted (Leal, Vrij, Mann, & Fisher, 2010).

The devil's advocate lie-detection tool was developed to detect truths and lies in expressing opinions. Interviewees are first asked an opinion-eliciting question that induces them to

argue in favor of their personal view ("What are your reasons for supporting the Americans in the war in Afghanistan?"). This is followed by a question that asks participants to argue against their personal view ("Playing devil's advocate, is there anything you can say against the involvement of the Americans in Afghanistan?").

People normally think more deeply about, and hence are likely to be more able to generate reasons that support rather than oppose their beliefs and opinions (Ajzen, 2001; Darley & Gross, 1983; Waenke & Bless, 2000). Therefore, truth tellers are likely to provide more information in their responses to the opinion-eliciting question than to the devil's advocate question. This pattern is unlikely to to be found in liars because, for them, the devil's advocate question is more compatible with their beliefs than is the opinion-eliciting question. In an experiment testing the devil's advocate approach (Leal et al., 2010), truth tellers' opinion-eliciting answers were longer than their devil's advocate answers. Also, observers judged that the truth tellers' opinion-eliciting answers sounded more immediate and plausible and revealed more emotional involvement than did their devil's advocate answers. No clear differences emerged in liars' answers to the two types of question. On the basis of these differences in speech content, 86% of truth tellers and 79% of liars were correctly classified.

The specific question approach: The strategic use of evidence. Guilty suspects (i.e., liars) and innocent suspects (i.e., truth tellers) enter police interviews in a different mental state (Granhag & Hartwig, 2008; Porter & Yuille, 1995). A guilty suspects will have unique knowledge about the crime, and this information, if it becomes known to the interviewer, will make it obvious that they are the perpetrator. A liar's main concern will be to ensure that the interviewer does not gain knowledge of their actions at the time of the crime. In contrast, innocent suspects face the opposite problem, fearing that the interviewer will not come to know what the suspect did at the time of the crime. Research has shown that these different mental states result in different strategies for liars and truth tellers (Colwell et al., 2006; Granhag & Strömwall, 2002; Granhag, Strömwall, & Hartwig, 2007; Hartwig et al., 2007; Strömwall et al., 2007). Guilty suspects are inclined to use avoidance strategies (e.g., in a free recall, avoid mentioning where they were at a certain place at a certain time) or denial strategies (e.g., denying to be at a certain place at a certain time when directly asked). In contrast, innocent suspects neither avoid nor escape but are forthcoming and tell the truth like it happened (Granhag & Hartwig, 2008).

The strategic-use-of-evidence (SUE) technique addresses how interviewers can consider these different strategies that guilty and innocent suspects use when they possess potentially incriminating information about a suspect (Granhag et al., 2007; Hartwig et al., 2006). Suppose that a man who left his briefcase in a bookstore on top of a box of stationery returns to find that his wallet has been stolen from the briefcase. Further suppose that the police found fingerprints on the briefcase that did not belong to the owner but did belong to another customer who had visited the bookshop. This makes the customer a suspect but not necessarily the culprit; perhaps the customer moved the briefcase to look in the box of stationery. In such circumstances, the police need to interview the suspect to find out the truth.

The first step of the SUE technique is to ask the suspect to describe his or her activities (in this example, to describe his or her activities in the bookshop) but not to reveal the fingerprint evidence. It is more likely that truth tellers will mention the briefcase than will liars. Truth tellers have nothing to hide and will recall what had happened, and this includes touching the briefcase; liars do not wish to associate themselves with the crime they have committed and thus distance themselves from the briefcase. However, not mentioning touching the briefcase still does not establish guilt, because truth tellers may simply have forgotten to mention this minor detail. In the second phase of the SUE technique, the questioning phase, the interviewer asks questions, including those involving the briefcase, without revealing the incriminating fingerprint evidence. There is a chance that a liar will deny having touched the briefcase and will thereby contradict the evidence known to the lie detector. A truth teller would be more likely to reveal that he or she had moved the briefcase. The third phase of the SUE technique is to reveal the evidence and ask the suspect to explain any contradictions between their account and the evidence. Here, it should be noted that some contradictions may be caused by factors other than deceit such as, for example, a truth teller discussing an event in the distant past may simply misremember some details. Hence, not every contradiction is a clear-cut sign of deception.

Hartwig et al. (2006) tested the SUE technique in their experiment, using the stolen wallet scenario previously mentioned. Swedish police trainees interviewed the mock suspects. Half of the interviewers were trained how to use the SUE technique before the experiment and were asked to use this technique in the subsequent interview. The other half of the interviewers did not receive training and were instructed to interview the suspects in the manner of their own choice. The untrained interviewers obtained a 56.1% accuracy rate, which is similar to that typically found in nonverbal and verbal deception-detection research (C.F. Bond & DePaulo, 2006; Vrij, 2008a). SUE-trained interviewers, however, obtained an 85.4% accuracy rate. It appeared that guilty suspects contradicted the evidence more than did innocent suspects, but more important is that they did so particularly when they were interviewed by SUE-trained interviewers.

The SUE technique differs from traditional police interviews in an important way. Traditionally, the police are inclined to present the evidence (e.g., "Your fingerprints have been found on the briefcase") at the beginning of the interview (Hartwig et al., 2006; Leo, 1996). As we mentioned earlier, the traditional police technique is limited because it gives the guilty suspects the opportunity to fabricate a story that is consistent with the evidence. The delayed disclosure of evidence approach has other benefits. First, it encourages interviewers to not show suspicion and enter the interview with an open

mind. Once people have made up their minds about the veracity of a message, they have the tendency to interpret additional information in such a way that it supports their decision (see the dangerous decisions theory previously discussed). As a result, after making up their minds, lie detectors run the risk of failing to notice further important information or of misinterpreting such information. Second, revealing suspicions may make truth tellers feel uncomfortable and this may result in the *Othello error*, the erroneous decision to interpret such nerves as a sign of guilt. Third, suspiciousness may also result in escape routes for liars. For example, it could result in them refusing to talk any longer (e.g., "Why should I speak to you? You don't believe me anyway!").

Imposing cognitive load. As discussed earlier, deception theories postulate that liars may be more nervous and may have to think harder than truth tellers. However, research has shown that liars often do not display cues of nervousness and cognitive load and that cues to deception are typically faint and unreliable. But can interviewers go one step further? Are there interview techniques that elicit and enhance differences in nervousness or cognitive load? Together with the National Research Council (2003), we do not think that questions can be asked that will necessarily raise more concern in liars than in truth tellers; thus none of the interventions that we will now discuss aim to raise concern in interviewees. But research has demonstrated that it is possible to enhance differences in cognitive load between truth tellers and liars (Vrij et al., 2008; Vrij, Mann, Leal, & Fisher, 2010), so this is the aim of the following interventions.

Lying can be more cognitively demanding than truth telling for six reasons. First, formulating a lie itself may be cognitively demanding. A liar needs to invent a story and must monitor his or her fabrication so that it is plausible and adheres to everything observers would know or might find out. In addition, liars must remember what they have said to whom in order to maintain consistency. Liars should also avoid making slips of the tongue, while refraining from providing new leads (Vrij, 2008a).

A second aspect of lying that adds to mental load is the fact that liars are typically less likely than truth tellers to take their credibility for granted (DePaulo et al., 2003; Kassin, 2005; Kassin, Appleby, & Torkildson-Perillo, 2010; Kassin & Gudjonsson, 2004; Kassin & Norwick, 2004). Truth tellers typically assume that their innocence shines through (Granhag et al., 2007; Kassin; Kassin et al., 2009; Kassin & Gudjonsson; Kassin & Norwick; Vrij, Mann, & Fisher, 2006b), which could be explained with the *illusion of transparency* (Gilovich, Savitsky, & Medvec, 1998), the belief that one's inner feelings will manifest themselves on the outside, and *belief in a just world* (Lerner, 1980), the belief that people will get what they deserve, and deserve what they get. Liars will be more inclined than truth tellers to monitor and control their demeanor in order to appear honest to the lie detector (DePaulo & Kirkendol, 1989), and such monitoring and controlling is cognitively demanding (Baumeister, 1998). For example, the guilty suspect may experience powerful emotions (e.g., fear, remorse, anger, or even excitement) that must be hidden or faked, and that may differ from those of the truth teller (Porter & ten Brinke, 2010). Consider a woman publicly pleading for the safe return of her partner who, in reality, she has murdered (see also Vrij & Mann, 2001b). She must monitor her body language and emotional expressions while keeping the details of the story straight. A high level of cognitive load accompanies high-stakes deception.

Third, because liars do not take credibility for granted, they may monitor *interviewers'* reactions more carefully in order to assess whether their lies appear to be successful (Buller & Burgoon, 1996; Schweitzer, Brodt, & Croson, 2002). Carefully monitoring an interviewer also requires cognitive resources.

Fourth, liars may be preoccupied by the task of reminding themselves to act and role play (DePaulo et al., 2003), which requires extra cognitive effort. Fifth, liars have to suppress the truth while they are lying, and this is also cognitively demanding (Spence et al., 2001). Last, while activation of the truth often happens automatically, activation of a lie is more intentional and deliberate, and thus it requires mental effort (Gilbert, 1991; Walczyk, Roper, Seemann, & Humphrey, 2003; Walczyk et al., 2005).

A lie detector could exploit the differential levels of cognitive load that truth tellers and liars experience, in order to discriminate more effectively between them. Liars who require more cognitive resources than truth tellers for the act of storytelling will have fewer cognitive resources left over than truth tellers will. This makes liars vulnerable, and so if cognitive demand is further raised—which could be achieved by making additional requests—liars may not be as good as truth tellers in coping with these additional requests.

One way to impose cognitive load on interviewees is by asking them to tell their stories in reverse order. This increases cognitive load because (a) it runs counter to the natural forward-order coding of sequentially occurring events (Gilbert & Fisher, 2006; Kahana, 1996) and (b) it disrupts reconstructing events from a schema (Geiselman & Callot, 1990). In one experiment, half of the liars and truth tellers were requested to recall their stories in reverse order, whereas no instruction was given to the other half of the participants (Vrij et al., 2008). More cues to deceit emerged in this reverse-order condition than in the control condition. More important is that observers who watched these videotaped interviews could distinguish between truths and lies better in the reverse-order condition than in the control condition. In the control condition, only 42% of the lies were correctly classified, well below what is found in a typical lie-detection experiment, suggesting that the lie-detection task in this experiment was particularly difficult. Yet, in the experimental condition, 60% of the lies were correctly classified, slightly more than what is typically found in lie-detection research.

Another way to increase cognitive load is by instructing interviewees to maintain eye contact with the interviewer (Beattie, 1981). When people have to concentrate on telling their stories, which is likely when they are requested to recall

what has happened, they are inclined to look every now and then away from their conversation partner (typically to a motionless point), because maintaining eye contact with the conversation partner is distracting (Doherty-Sneddon, Bruce, Bonner, Longbotham, & Doyle, 2002; Doherty-Sneddon & Phelps, 2005; Glenberg, Schroeder, & Robertson, 1998). When interviewees are instructed to maintain eye contact continuously, their concentration on telling their stories is therefore likely to be hampered, and, because lying is more mentally taxing than truth telling, this should impair the storytelling of liars more than the storytelling of truth tellers. In one experiment, half of the liars and truth tellers were requested to maintain eye contact with the interviewer continuously throughout the interview, whereas no instruction was given to the other half of the participants (Vrij, Mann, Leal, & Fisher, 2010). It was again found that more cues to deceit emerged in the eye-contact condition than in the control condition and that observers who watched these videotaped interviews could discriminate between truths and lies only in the eye-contact condition.

An experiment with children reveals a third type of additional request that can be made to increase a liar's cognitive load: asking event-irrelevant questions (Quas, Davis, Goodman, & Myers, 2007). Children played individually with a male confederate who touched each child twice on their stomach, nose, and neck. In the subsequent interview, children were asked to tell the truth or lie when asked questions about the touching. They also were asked a series of questions about the event that were unrelated to body touch and were asked to answer those questions truthfully. The children who lied about the body touch answered these unrelated questions less accurately than did the children who told the truth about the body touch. Quas et al. argued that remembering and rehearsing the lie required cognitive resources and that by devoting their resources to the lie, children had difficulty in conducting an adequate memory search for other event details.

Future Research Directions

Although the nonverbal and verbal deception-detection literature is extensive, several important issues still remain to be addressed. We acknowledge four issues that we believe are fruitful and important avenues for future research. First, although much research has aimed at discriminating between truths and lies about past actions, virtually no research has been conducted on distinguishing between truths and lies about future actions (intentions). This is remarkable considering the frequency and importance of situations calling for assessments of whether a person is lying or truth telling about his or her intentions (e.g., stated reasons for crossing a border, for example). Consider the would-be 911 terrorists, smiling and chatting politely with airport staff while perhaps covertly feeling intense hatred and contempt toward their intended targets, as well as fear of discovery and/or death. Is it possible to identify such individuals by their behavior or responses to specific questions? The societal value of being able to detect planned but not-yet-committed illegal actions (criminal intentions) is thus obvious (Granhag, 2010).

Deception research about intentions has commenced with the publication of three experimental studies (Granhag & Knieps, in press; Vrij, Granhag, Mann, & Leal, in press; Vrij, Leal, Mann, et al., in press). The pattern that emerges from these experiments is that deceptive intentions are associated with different cues to deceit than are deceptive descriptions of past activities. For example, research on past activities has shown that typically liars are less detailed than truth tellers (DePaulo et al., 2003; Vrij, 2005, 2008a), whereas no difference in detail emerged in any of the deceptive-intention experiments so far. One aspect that often makes truth tellers' stories about past activities more detailed than liars' stories is that there is a wealth of perceptual details that truth tellers have experienced during these past activities that they can recall (if they still remember them). In contrast, when discussing their intentions about a forthcoming activity, truth tellers have not yet experienced anything, and this restricts the amount of detail in their recall of intentions.

Some differences between truthful and deceptive intentions emerged. First, truthful intentions sounded more plausible than did deceptive intentions (Vrij, Granhag, Mann, & Leal, in press; Vrij, Leal, Mann, et al., in press), and truthful and deceptive intentions were associated with different mental images (Granhag & Kniep, in press). Participants who told the truth about their intentions agreed more frequently that planning their future actions evoked mental images than did participants who lied about their intentions. In addition, liars who claimed to have activated a mental image during the planning phase provided verbal descriptions of the most dominant mental image that were less rich in detail than those of the truth tellers. Those findings align with the concept of episodic future thought. In brief, episodic future thought represents the ability to mentally preexperience a one-time personal event that may occur in the future (Schacter & Addis, 2007). People who make up a plan for a future event that they intend to execute seem to activate a more concrete (detailed) mental image of the upcoming scenario than do those who adopt a plan that they do not intend to execute (Watanabe, 2005).

A second line of research that needs greater attention is work with real populations, such as actual suspects, and high-stakes lies. In fact, only three studies of high-stakes lies with actual suspects have been conducted (Mann et al., 2002; Vrij & Mann, 2001a, 2001b). Porter and ten Brinke (2010) argue that there may be qualitative and quantitative variations in the behavioral manifestations of lies of minor consequence versus those of major consequence. Although high-stakes lies may be harder for liars to tell, their behavioral signs are neither obvious (i.e., police perform just above chance when trying to identify them; Vrij & Mann, 2001b) and may simply not be more extreme than those of lower-stakes lies.

A third line of research that merits attention is lying by networks. Most deception research addresses individual truth tellers and liars, but criminals often act in pairs or larger groups. Research could focus on the development of interview tools

that can successfully discriminate between pairs of truth tellers and pairs of liars. Probably the dominant interview strategy to date is to interview each member of the group individually and compare the answers they give. If the members give consistent answers, they are considered truth tellers; if they give contradicting answers, they are considered liars. This strategy is limited, because it appears to ignore the fact that liars tend to prepare their alibis together, and therefore they are likely to give the same answers when asked about these alibis. The strategy works, however, if questions are asked that the liars have not anticipated, because in that case they cannot give their prepared answers (Vrij et al., 2009). Thus, examining contradictions could work, but only with answers to unanticipated questions. There is no evidence that professionals make this crucial distinction between anticipated and unanticipated questions when they interview multiple suspects.

A fourth line of fruitful and important research is examining the strategies used by truth tellers and liars when they are interviewed. As we have argued here, effective lie-detection interview techniques take advantage of the distinctive psychological processes and requirements of truth tellers and liars. To design such interview strategies, we need further insight into truth tellers' and liars' strategies through research. For example, research has shown that verbal cues are typically more diagnostic cues to deceit than are nonverbal cues (DePaulo et al., 2003; Vrij, 2008a, 2008b), and truth tellers' and liars' strategies can explain this. In one study, truth tellers and liars were found to use different verbal strategies (Vrij, Mann, Leal, & Granhag, 2010). Truth tellers were mainly concerned with telling what had happened. In contrast, liars were preparing their answers to possible questions. Liars further decided not to give too much detail, because providing details increases the chance of saying something that the interviewer knows to be untrue. The result of these different verbal strategies is that truth tellers' stories are likely to be more detailed than those of liars; research by DePaulo et al. (2003) and Vrij (2008a) supports this idea. Although truth tellers and liars in these studies did use different verbal strategies, they used the same nonverbal strategies. Both truth tellers and liars believed that signs of nervousness would appear suspicious. They therefore decided that they would try to suppress displaying signs of nervousness during the interview. The fact that truth tellers and liars employ different verbal strategies but the same nonverbal strategies (a finding also obtained by Hartwig, Granhag, Strömwall, & Doering, 2010) may explain, in part, why verbal cues to deceit are often more diagnostic than are nonverbal cues to deceit.

Conclusion

We have presented an overview of pitfalls and opportunities in nonverbal and verbal lie detection. We presented 16 pitfalls and clustered them into three categories: (a) a lack of motivation to detect lies, (b) difficulties associated with lie detection, and (c) common errors made by lie detectors. We believe that the most important point to take home is that nonverbal and verbal cues to deception are ordinarily faint and unreliable. This makes lie detection a difficult task, as there is no nonverbal or verbal cue that lie detectors can truly rely upon.

We also discussed 11 guidelines to improve lie detection. First, we presented 5 guidelines aimed at avoiding common errors made in nonverbal and verbal lie detection. This has been the focus of research for a considerable period of time. We then discussed 6 guidelines aimed at creating more cues and more blatant and reliable cues to deception by exploiting truth tellers' and liars' distinctive psychological states. This has been the focus of recent research. We believe that the success of the traditional methods to improve lie detection is seriously hampered by the fact that cues are typically faint and unreliable. The recently introduced methods attempt to tackle exactly this problem, and, as we have demonstrated, are doing so with success. We encourage lie detectors to become actively engaged in exploiting truth tellers' and liars' different mental processes. This should not be restricted to police–suspect interviews, the topic of investigation in many deception experiments. It could equally be used in a variety of settings, including an intelligence context for the identification and apprehension of individuals with criminal intent. It may even be used for detecting lies told in the courtroom. We encourage researchers to focus their efforts on this line of innovative and promising lie-detection research.

Endnotes

1. Not all probing questions facilitate lie detection. In many earlier studies examining the effect of questioning, probes such as "I don't understand this, could you please explain this to me?" (neutral probes); "I do believe you, but I don't understand this. How is it possible that...?" (positive probes); or "I don't believe you, are you trying to fool me?" (negative probes) were used. Intuitively, one might think that such probes make truth detection and lie detection easier: The liar is forced to continue to speak and give more information; and the more liars speak and the more information they give, the greater the possibility that they will make mistakes and give their lies away, either via verbal cues (by contradicting themselves or by saying something which an observer knows is incorrect) or via nonverbal cues. However, several studies have shown that these types of probing do not increase accuracy but tend to lead to judging the other as being truthful (G.D. Bond, Malloy, Thompson, Arias, & Nunn, 2004; Buller, Comstock, Aune, & Strzyzewski, 1989; Buller, Strzyzewski, & Comstock, 1991; Levine & McCornack, 2001; Stiff & Miller, 1986). This is called the *probing heuristic* (Levine, Park, & McCornack, 1999). The type of probing (negative, neutral, or positive) is irrelevant; all types of probing yield the same effect and benefit liars. In the "Exploiting the Different Mental Processes of Truth Tellers and Liars" section of this review, we discuss successful probing questions.
2. Note that when people overwhelmingly say that liars avert their gaze, it does not mean that they always rely on gaze aversion when they attempt to detect deceit. For example, Vrij (1993) correlated the behaviors displayed by the videotaped liars and truth tellers (e.g., gaze behavior, smiling, different types of movements, stutters) with the veracity judgments made by the police detectives

who observed these videotapes. The gaze patterns displayed by the liars and truth tellers did not predict the police detectives' veracity judgments in this particular study, whereas smiling (people who smiled less were perceived as more suspicious) and movements (people who moved their arms and hands more were perceived as more suspicious) did. In a meta-analysis of such studies, Hartwig and Bond (2010) found a correlation of $r = .27$ between averting gaze and veracity judgements (people who avert their gaze are perceived as more suspicious). Although this correlation was significant, it was somewhat lower than some other correlations. The cues that had the strongest relation with veracity judgments were incompetence ($r = -.54$) and ambivalence ($r = .51$). People who appear incompetent and/or ambivalent are judged as deceptive.

3. There are many interrogation manuals, and they are highly similar to each other (Vrij & Granhag, 2007). We mainly focus on the Inbau et al. (2001) manual, because this manual is commonly used by police and military interrogators and hence is highly influential (Gudjonsson, 2003).
4. Throughout the years, the Ekman group in particular has claimed that individual differences in the ability to detect deceit exist. They first reported that some groups of professionals (e.g., The Secret Service) are better lie detectors than other groups (Ekman & O'Sullivan, 1991; Ekman, O'Sullivan, & Frank, 1999). Later they reported that they had identified some individuals with extraordinarily good skills in lie detection, the so-called wizards (O'Sullivan & Ekman, 2004). Charles F. Bond has challenged these findings, arguing that individual differences are minute (Bond & DePaulo, 2008). Regarding the group differences, C.F. Bond (2008) noticed that a draft manuscript from Ekman et al.'s 1999 article, circulated in 1997, differed from the final 1999 article and that not all the findings reported in the 1997 draft were included in the 1999 article. Because the findings in 1999 were more in alignment with Ekman et al.'s argument about the superiority of certain groups in lie detection than the findings in the 1997 draft, C.F. Bond (2008) suspected manipulation and believed that Ekman and colleagues avoided reporting the findings that went against their general conclusion. Ekman, O'Sullivan, and Frank (2008) denied manipulation. They reported that after 1997, they tested additional groups of participants but that these new groups did not complete all the lie-detection tests that the earlier groups had completed. In their 1999 article, they only reported the results for the lie-detection tests that were completed by all the groups. Regarding their findings, Bond and Uysal (2007) reasoned that the number of wizards that were identified was so low (15 out of 13,000 people who were tested) that they could have emerged as wizards just by chance. However, O'Sullivan (2007) argued that subsequent follow-up testing has demonstrated that these wizards were true wizards. More important for this article is whether wizards use clearly identifiable strategies. If so, it would mean that others could learn from them. The Ekman group has not published detailed data about the strategies used by their wizards to date, but G.D. Bond (2009) has. In his wizard project, G.D. Bond started with 234 lie detectors and identified two wizards. Via eye-tracking equipment he determined the locations the two wizards looked at when making their veracity decisions. The two experts used different strategies: One wizard looked more at the face area, whereas the other looked more at the arm/torso area. In summary, if wizards exist, it is so far unclear what makes them wizards. O'Sullivan and colleagues further claimed that truth and lie detection becomes easier when there is more at stake for the truth tellers and liars (O'Sullivan, 2008; O'Sullivan, Frank, Hurley, & Tiwana, in press). This claim has been supported by experimental research (DePaulo, Blank, Swaim, & Hairfield, 1992; DePaulo, Kirkendol, Tang, & O'Brien, 1988; DePaulo, Lanier, & Davis, 1983; DePaulo, LeMay, & Epstein, 1991; DePaulo, Stone, & Lassiter, 1985; Lane & DePaulo, 1999; Vrij, 2000; Vrij, Harden, Terry, Edward, & Bull, 2001).
5. Many of these guidelines require interviewees to talk. We believe that interviewees are generally willing to talk even in situations in which such willingness may be less expected, such as in police interviews. In their analysis of 1,067 audiotaped police interviews, Moston, Stephenson, and Williamson (1993) found that only 5% of suspects remained silent.

Acknowledgments

The authors are very grateful to Bella M. DePaulo for her constructive comments on an earlier draft of this article.

References

Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology, 52*, 27–58.

Albrechtsen, J.S., Meissner, C.A., & Susa, K.J. (2009). Can intuition improve deception detection performance? *Journal of Experimental Social Psychology, 45*, 1052–1055.

Allwood, C.M., & Granhag, P.A. (1999). Feelings of confidence and the realism of confidence judgments in everyday life. In P. Juslin, & H. Montgomery (Eds.), *Judgment and decision making: Neo-Brunswikian and process-tracing approaches* (pp. 123–146). Mahwah, NJ: Erlbaum.

Ambady, N., Bernieri, F.J., & Richeson, J.A. (2000). Toward a histology of social behaviour: Judgmental accuracy from thin slices of the behavioural stream. *Advances in Experimental Social Psychology, 32*, 201–271.

Anderson, C.A., Lepper, M.R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39*, 1037–1049.

Anderson, D.E., Ansfield, M.E., & DePaulo, B.M. (1999). Love's best habit: Deception in the context of relationships. In P. Philippot, R.S. Feldman, & E.J. Coats (Eds.), *The social context of nonverbal behavior* (pp. 372–409). Cambridge, England: Cambridge University Press.

Anderson, D.E., DePaulo, B.M., Ansfield, M.E., Tickle, J.J., & Green, E. (1999). Beliefs about cues to deception: Mindless stereotypes or untapped wisdom? *Journal of Nonverbal Behaviour, 23*, 67–89.

Arntzen, F. (1970). *Psychologie der Zeugenaussage [Psychology of testimony]*. Göttingen, Germany: Hogrefe.

Arntzen, F. (1982). Die Situation der forensischen Aussagenpsychologie in der Bundesrepublik Deutschland. In A. Trankell (Ed.), *Reconstructing the past: The role of psychologists in criminal trials* (pp. 107–120). Deventer, The Netherlands: Kluwer.

Arntzen, F. (1983). *Psychologie der Zeugenaussage: Systematik der Glaubwürdigkeitsmerkmale* [Psychology of testimony: A systematic approach of credibility assessment]. München, Germany: Beck.

Backbier, E., Hoogstraten, J., & Meerum Terwogt-Kouwenhoven, K. (1997). Situational determinants of the acceptability of telling lies. *Journal of Applied Social Psychology, 27*, 1048–1062.

Bageley, J., & Manelis, L. (1979). Effect of awareness on an indicator of cognitive load. *Perceptual and Motor Skills, 49*, 591–594.

Baldry, A.C., & Winkel, F.W. (1998). Perceptions of the credibility and eventual value of victim and suspect statements in interviews. In J. Boros, I. Munnich, & M. Szegedi (Eds.), *Psychology and criminal justice: International review of theory and practice* (pp. 74–82). Berlin, Germany: de Gruyter.

Baldry, A.C., Winkel, F.W., & Enthoven, D.S. (1997). Paralinguistic and nonverbal triggers of biased credibility assessments of rape victims in Dutch police officers: An experimental study of "nonevidentiary" bias. In S. Redondo, V. Garrido, J. Perze, & R. Barbaret (Eds.), *Advances in psychology and law* (pp. 163–174). Berlin, Germany: de Gruyter.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*, 269–278.

Barrett, E.C. (2005). Psychological research and police investigations: Does the research meet the needs? In L. Alison (Ed.), *The forensic psychologist's casebook* (pp. 47–67). Devon, England: Willan.

Baumeister, R.F. (1998). The self. In D.T. Gilbert, S.T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., pp. 680–740). Boston: McGraw-Hill.

Bavelas, J.B., & Chovil, N. (2006). Nonverbal and verbal communication: Hand gestures and facial displays as part of language use in face-to-face dialogue. In V. Manusov, & M.L. Patterson (Eds.), *The SAGE handbook of nonverbal communication* (pp. 97–115). Thousand Oaks, CA: Sage.

Bavelas, J.B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin, 21*, 394–405.

Bavelas, J.B., & Gerwing, J. (2007). Conversational hand gestures and facial displays in face-to-face dialogue. In K. Fiedler (Ed.), *Frontiers of social psychology: Social communication* (pp. 283–307). New York: Psychology Press.

Beattie, G.W. (1981). A further investigation of the cognitive interference hypothesis of gaze patterns during conversation. *British Journal of Social Psychology, 20*, 243–248.

Bell, K.L., & DePaulo, B.M. (1996). Liking and lying. *Basic and Applied Social Psychology, 18*, 243–266.

Blair, J.P., & Kooi, B. (2004). The gap between training and research in the detection of deception. *International Journal of Police Science and Management, 6*, 77–83.

Bok, S. (1989). *Lying: Moral choice in public and private life*. New York: Random House.

Bollingmo, G., Wessel, E., Sandvold, Y., Eilertsen, D.E., & Magnussen, S. (2009). The effect of biased and non-biased information on judgements of witness credibility. *Psychology, Crime & Law, 15*, 61–71.

Bond, C.F. (2008). A few can catch a liar, sometimes. *Applied Cognitive Psychology, 22*, 1298–1300.

Bond, C.F., & DePaulo, B.M. (2006). Accuracy of deception judgements. *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.

Bond, C.F., & Fahey, W.E. (1987). False suspicion and the misperception of deceit. *British Journal of Social Psychology, 26*, 41–46.

Bond, C.F., & Uysal, A. (2007). On lie detection "wizards." *Law and Human Behavior, 31*, 109–115.

Bond, G.D. (2009). Deception detection expertise. *Law and Human Behavior, 32*, 339–351.

Bond, G.D., Malloy, D.M., Thompson, L.A., Arias, E.A., & Nunn, S. (2004). Post-probe decision making in a prison context. *Communication Monographs, 71*, 269–285.

Bothwell, R., & Jalil, M. (1992). The credibility of nervous witnesses. *Journal of Social Behavior and Personality, 7*, 581–586.

Brant, C. (1993). Communication patterns in Indians: Verbal and nonverbal. *Sexual Abuse: A Journal of Research and Treatment, 6*, 259–269.

Bull, P. (2009). Detecting deceit: Current issues. In T. Williamson, B. Milne, & S.P. Savage (Eds.), *International developments in investigative interviewing* (pp. 190–206). Cullompton, England: Willan.

Bull, R., & McAlpine, S. (1998). Facial appearance and criminality. In A. Memon, A. Vrij, & R. Bull (Eds.), *Psychology and law: Truthfulness, accuracy and credibility* (pp. 59–76). London: McGraw-Hill.

Buller, D.B., & Aune, R.K. (1988). The effects of vocalics and nonverbal sensitivity on compliance: A speech accommodation theory explanation. *Human Communication Research, 14*, 301–332.

Buller, D.B., & Burgoon, J.K. (1996). Interpersonal deception theory. *Communication Theory, 6*, 203–242.

Buller, D.B., Comstock, J., Aune, R.K., & Strzyzewski, K.D. (1989). The effect of probing on deceivers and truthtellers. *Journal of Nonverbal Behavior, 13*, 155–170.

Buller, D.B., Strzyzewski, K.D., & Comstock, J. (1991). Interpersonal deception: I. Deceivers' reactions to receivers' suspicions and probing. *Communication Monographs, 58*, 1–24.

Burgess, A.W. (1985). *Rape and sexual assault: A research book*. London: Garland.

Burgess, A.W., & Homstrom, L.L. (1974). *Rape: Victims of crisis*. Bowie, MD: Brady.

Burgoon, J.K., Blair, J.P., & Strom, R.E. (2008). Cognitive biases and nonverbal cue availability in deception detection. *Human Communication Research, 34*, 572–599.

Burgoon, J.K., & Buller, D.B. (1994). Interpersonal deception: III. Effects of deceit on perceived communication and nonverbal dynamics. *Journal of Nonverbal Behavior, 18*, 155–184.

Burgoon, J.K., Buller, D.B., Floyd, K., & Grandpre, J. (1996). Deceptive realities: Sender, receiver, and observer perspectives in deceptive conversations. *Communication Research, 23*, 724–748.

Burgoon, J.K., Buller, D.B., White, C.H., Afifi, W., & Buslig, A.L.S. (1999). The role of conversation involvement in deceptive interpersonal interactions. *Personality and Social Psychology Bulletin, 25*, 669–685.

Caso, L., Maricchiolo, F., Bonaiuto, M., Vrij, A., & Mann, S. (2006). The impact of deception and suspicion on different hand movements. *Journal of Nonverbal Behavior, 30*, 1–19.

Caso, L., Vrij, A., Mann, S., & de Leo, G. (2006). Deceptive responses: The impact of verbal and nonverbal countermeasures. *Legal and Criminological Psychology, 11*, 99–111.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Colwell, L.H., Miller, H.A., Lyons, P.M., & Miller, R.S. (2006). The training of law enforcement officers in detecting deception: A survey of current practices and suggestions for improving accuracy. *Police Quarterly, 9*, 275–290.

Dalton, A.L., & Daneman, M. (2006). Social suggestibility to central and peripheral misinformation. *Memory, 14*, 486–501.

Darley, J.M., & Gross, P.H. (1983). A hypothesis-confirming bias in labelling effects. *Journal of Personality and Social Psychology, 44*, 20–33.

Davies, G.M. (1991). Research on children's testimony: Implications for interviewing practice. In C.R. Hollin, & K. Howells (Eds.), *Clinical approaches to sex offenders and their victims* (pp. 177–191). New York: Wiley.

Davis, M., & Hadiks, D. (1995). Demeanor and credibility. *Semiotica, 106*, 5–54.

DePaulo, B.M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin, 111*, 203–243.

DePaulo, B.M. (2004). The many faces of lies. In A.G. Miller (Ed.), *The social psychology of good and evil* (pp. 303–236). New York: Guilford Press.

DePaulo, B.M., Blank, A.L., Swaim, G.W., & Hairfield, J.G. (1992). Expressiveness and expressive control. *Personality and Social Psychology Bulletin, 18*, 276–285.

DePaulo, B.M., Charlton, K., Cooper, H., Lindsay, J. L., & Muhlenbruck, L. (1997). The accuracy–confidence correlation in the detection of deception. *Personality and Social Psychology Review, 1*, 346–357.

DePaulo, B.M., Epstein, J.A., & LeMay, C.S. (1990). Responses of the socially anxious to the prospect of interpersonal evaluation. *Journal of Personality, 58*, 623–640.

DePaulo, B.M., & Friedman, H.S. (1998). Nonverbal communication. In D.T. Gilbert, S.T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 3–40). Boston: McGraw-Hill.

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., & Kirkendol, S.E. (1989). The motivational impairment effect in the communication of deception. In J.C. Yuille (Ed.), *Credibility assessment* (pp. 51–70). Dordrecht, The Netherlands: Kluwer.

DePaulo, B.M., Kirkendol, S.E., Tang, J., & O'Brien, T.P. (1988). The motivational impairment effect in the communication of deception: Replications and extensions. *Journal of Nonverbal Behavior, 12*, 177–202.

DePaulo, B.M., Lanier, K., & Davis, T. (1983). Detecting deceit of the motivated liar. *Journal of Personality and Social Psychology, 45*, 1096–1103.

DePaulo, B.M., LeMay, C.S., & Epstein, J.A. (1991). Effects of importance of success and expectations for success on effectiveness at deceiving. *Personality and Social Psychology Bulletin, 17*, 14–24.

DePaulo, B.M., Lindsay, J.L., Malone, B.E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74–118.

DePaulo, B.M., & Pfeifer, R.L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology, 16*, 249–267.

DePaulo, B.M., Rosenthal, R., Eisenstat, R.A., Rogers, P.L., & Finkelstein, S. (1978). Decoding discrepant nonverbal cues. *Journal of Personality and Social Psychology, 36*, 313–323.

DePaulo, B.M., Stone, J.I., & Lassiter, G.D. (1985). Telling ingratiating lies: Effects of target sex and target attractiveness on verbal and nonverbal deceptive success. *Journal of Personality and Social Psychology, 48*, 1191–1203.

Doherty-Sneddon, G., Bruce, V., Bonner, L., Longbotham, S., & Doyle, C. (2002). Development of gaze aversion as disengagement of visual information. *Developmental Psychology, 38*, 438–445.

Doherty-Sneddon, G., & Phelps, F.G. (2005). Gaze aversion: A response to cognitive or social difficulty? *Memory and Cognition, 33*, 727–733.

Dumas, R., & Testé, B. (2006). The influence of criminal facial stereotypes on juridic judgments. *Swiss Journal of Psychology, 65*, 237–244.

Ekman, P. (1997). Deception, lying, and demeanor. In D.F. Halpern, & A.E. Voiskounsky (Eds.), *States of mind: American and post-Soviet perspectives on contemporary issues in psychology* (pp. 93–105). New York: Oxford University Press.

Ekman, P. (2001). *Telling lies: Clues to deceit in the marketplace, politics and marriage*. New York: Norton. (Original work published 1985).

Ekman, P., & Friesen, W.V. (1969). Nonverbal leakage and clues to deception. *Psychiatry, 32*, 88–106.

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist, 46*, 913–920.

Ekman, P., & O'Sullivan, M. (2006). From flawed self-assessment to blatant whoppers: The utility of voluntary and involuntary behavior in detecting deception. *Behavioural Sciences & the Law, 24*, 673–686.

Ekman, P., O'Sullivan, M., & Frank, M.G. (1999). A few can catch a liar. *Psychological Science, 10*, 263–266.

Ekman, P., O'Sullivan, M., & Frank, M.G. (2008). Scoring and reporting: A response to Bond (2008). *Applied Cognitive Psychology, 22*, 1315–1317.

Ekman, P., O'Sullivan, M., Friesen, W.V., & Scherer, K. (1991). Face, voice, and body in detecting deceit. *Journal of Nonverbal Behavior, 15*, 125–135.

Elaad, E. (2003). Effects of feedback on the overestimated capacity to detect lies and the underestimated ability to tell lies. *Applied Cognitive Psychology, 17*, 349–363.

Feeley, T.H., & Young, M.J. (2000). The effects of cognitive capacity on beliefs about deceptive communication. *Communication Quarterly, 48*, 101–119.

Feldman, S.S., & Cauffman, E. (1999). Sexual betrayal among late adolescents: Perspectives of the perpetrator and the aggrieved. *Journal of Youth and Adolescence, 28*, 235–258.

Fiedler, K., & Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global heuristics. *Human Communication Research, 20*, 199–223.

Fisher, R.P., Brennan, K.H., & McCauley, M.R. (2002). The cognitive interview method to enhance eyewitness recall. In M.L. Eisen, J.A. Quas, & G.S. Goodman (Eds.), *Memory and suggestibility in the forensic interview* (pp. 265–286). Mahwah, NJ: Erlbaum.

Fisher, R.P., Vrij, A., & Leins, D.A. (in press). Inconsistency as a predictor of memory inaccuracy and lying. In B.S. Cooper, D. Griesel, & M. Ternes (Eds.), *Applied issues in investigative interviewing, eyewitness memory, and credibility assessment* New York, NY: Springer.

Ford, E.B. (2006). Lie detection: Historical, neuropsychiatric and legal dimensions. *International Journal of Law and Psychiatry, 29*, 159–177.

Fowler, K.A., Lilienfeld, S.O., & Patrick, C.J. (2009). Detecting psychopathy from thin slices of behavior. *Psychological Assessment, 21*, 68–78.

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.

Frank, M.G., & Feeley, T.H. (2003). To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research, 31*, 58–75.

Frank, M.G., Yarbrough, J.D., & Ekman, P. (2006). Investigative interviewing and the detection of deception. In T. Williamson (Ed.), *Investigative interviewing: Rights, research and regulation* (pp. 229–255). Cullompton, England: Willan.

Freedman, N. (1972). The analysis of movement behavior during the clinical interview. In A.R. Siegman, & B. Pope (Eds.), *Studies in dyadic communication* (pp. 153–175). Elmsford, NY: Pergamon.

Ganis, G., Kosslyn, S.M., Stose, S., Thompson, W.L., & Yurgelun-Todd, D.A. (2003). Neural correlates of different types of deception: An fMRI investigation. *Cerebral Cortex, 13*, 830–836.

Garrido, E., Masip, J., & Herrero, C. (2004). Police officers' credibility judgements: Accuracy and estimated ability. *International Journal of Psychology, 39*, 254–275.

Geiselman, R.E., Callot, R. (1990). Reverse versus forward recall of script-based texts. *Journal of Applied Cognitive Psychology, 4*, 141–144.

Gigerenzer, G., Todd, P.M., & the ABC Research Group. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Gilbert, D.T. (1991). How mental systems believe. *American Psychologist, 46*, 107–119.

Gilbert, J.A.B., & Fisher, R.P. (2006). The effects of varied retrieval cues on reminiscence in eyewitness memory. *Applied Cognitive Psychology, 20*, 723–739.

Gilovich, T., Savitsky, K., & Medvec, V.H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology, 75*, 332–346.

Glenberg, A.M., Schroeder, J.L., & Robertson, D.A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition, 26*, 651–658.

The Global Deception Team. (2006). A world of lies. *Journal of Cross-Cultural Psychology, 37*, 60–74.

Goodman, G.S., Rudy, L., Bottoms, B., & Aman, C. (1990). Children's concerns and memory: Issues of ecological validity in the study of children's eyewitness testimony. In R. Fivush, & J. Hudson (Eds.), *Knowing and remembering in young children* (pp. 249–284). New York: Cambridge University Press.

Granhag, P.A. (2010). On the psycho-legal study of true and false intentions: Dangerous waters and some stepping stones. *The Open Criminology Journal, 3*, 37–43.

Granhag, P.A., Andersson, L.O., Strömwall, L.A., & Hartwig, M. (2004). Imprisoned knowledge: Criminals' beliefs about deception. *Legal and Criminological Psychology, 9*, 103–119.

Granhag, P.A., & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind-reading. *Psychology, Crime & Law, 14*, 189–200.

Granhag, P.A., & Knieps, M. (in press). Episodic future thought: Illuminating the trademarks of true and false intent. *Applied Cognitive Psychology.*

Granhag, P.A., & Strömwall, L.A. (1999). Repeated interrogations: Stretching the deception detection paradigm. *Expert Evidence: The International Journal of Behavioural Sciences in Legal Contexts, 7*, 163–174.

Granhag, P.A., & Strömwall, L.A. (2000a). "Let's go over this again...": Effects of repeated interrogations on deception detection performance. In A. Czeredrecka, T. Jaskiewicz-Obydzinska, & J. Wojcikiewicz (Eds.), *Forensic psychology and law: Traditional questions and new ideas* (pp. 191–196). Kraków, Poland: Institute of Forensic Research.

Granhag, P.A., & Strömwall, L.A. (2000b). Deception detection: Examining the consistency heuristic. In C.M. Breur, M.M. Kommer, J.F. Nijboer, & J.M. Reintjes (Eds.), *New trends in criminal investigation and evidence* (Vol. 2, pp. 309–321). Antwerpen, Belgium: Intresentia.

Granhag, P.A., & Strömwall, L.A. (2001). Deception detection: Interrogators' and observers' decoding of consecutive statements. *The Journal of Psychology, 135*, 603–620.

Granhag, P.A., & Strömwall, L.A. (2002). Repeated interrogations: Verbal and nonverbal cues to deception. *Applied Cognitive Psychology, 16*, 243–257.

Granhag, P.A., & Strömwall, L.A. (2004). *The detection of deception in forensic contexts.* Cambridge, UK: Cambridge University Press.

Granhag, P.A., Strömwall, L.A., & Hartwig, M. (2005). Granting asylum or not? Migration board personnel's beliefs about deception. *Journal of Ethnic and Migration Studies, 31*, 29–50.

Granhag, P.A., Strömwall, L.A., & Hartwig, M. (2007). The SUE technique: The way to interview to detect deception. *Forensic Update, 88*, 25–29.

Granhag, P.A., Strömwall, L.A., & Jonsson, A.C. (2003). Partners in crime: How liars in collusion betray themselves. *Journal of Applied Social Psychology, 33*, 848–868.

Granhag, P.A., & Vrij, A. (2010). Interviewing to detect deception. In P.A. Granhag (Ed.), *Forensic psychology in context: Nordic and*

international approaches (pp. 75–93). Cullompton, England: Willan.

Greely, H., & Illes, J. (2007). Neuroscience-based lie detection: The urgent need for regulation. *American Journal of Law and Medicine, 33*, 377–431.

Greuel, L. (1992). Police officers' beliefs about cues associated with deception in rape cases. In F. Lösel, D. Bender, & T. Bliesener (Eds.), *Psychology and law: International perspectives* (pp. 234–239). Berlin, Germany: de Gruyter.

Gudjonsson, G.H. (2003). *The psychology of interrogations and confessions*. Chichester, England: Wiley.

Hale, J.L., & Stiff, J.B. (1990). Nonverbal primacy in veracity judgments. *Communication Reports, 3*, 75–83.

Hare, R.D. (2006). Psychopathy: A clinical and forensic overview. *Psychiatric Clinics of North America, 29*, 709–724.

Hartwig, M., & Bond, C.F. (2010). *Why do lie-catchers fail? A Brunswikian lens model of human lie judgments*. Manuscript submitted for publication.

Hartwig, M., Granhag, P.A., & Strömwall, L. (2007). Guilty and innocent suspects' strategies during interrogations. *Psychology, Crime, & Law, 13*, 213–227.

Hartwig, M., Granhag, P.A., Strömwall, L., & Doering, N. (2010). Impression and information management: On the strategic self-regulation of innocent and guilty suspects [Special issue]. *The Open Criminology Journal, 3*, 10–16.

Hartwig, M., Granhag, P.A., Strömwall, L., & Kronkvist, O. (2006). Strategic use of evidence during police interrogations: When training to detect deception works. *Law and Human Behavior, 30*, 603–619.

Hartwig, M., Granhag, P.A., Strömwall, L., & Vrij, A. (2005). Detecting deception via strategic disclosure of evidence. *Law and Human Behavior, 29*, 469–484.

Henig, R.M. (2006, February 5). Looking for the lie. *The New York Times*. Retrieved July 12, 2007, from http://www.nytimes.com/2006/02/05/magazine/05lying.html

Hirsch, A.R., & Wolf, C.J. (2001). Practical methods for detecting mendacity: A case study. *Journal of the American Academy of Psychiatry and the Law, 29*, 438–444.

Hocking, J.E., & Leathers, D.G. (1980). Nonverbal indicators of deception: A new theoretical perspective. *Communication Monographs, 47*, 119–131.

Horvath, F., Jayne, B., & Buckley, J. (1994). Differentiation of truthful and deceptive criminal suspects in behavioral analysis interviews. *Journal of Forensic Sciences, 39*, 793–807.

Inbau, F.E., Reid, J.E., Buckley, J.P., & Jayne, B.C. (2001). *Criminal interrogation and confessions*. (4th ed.). Gaithersburg, MD: Aspen.

Joffe, R., & Yuille, J.C. (1992, May). *Criteria-based content analysis: An experimental investigation*. Paper presented at the North Atlantic Treaty Organization Advanced Study Institute on the Child Witness in Context: Cognitive, Social and Legal Perspectives, Lucca, Italy.

Johnson, R.R. (2006a). Confounding influences on police detection of suspiciousness. *Journal of Criminal Justice, 34*, 435–442.

Johnson, R.R. (2006b). Race and police reliance on suspicious nonverbal cues. *Policing: An International Journal of Police Strategies & Management, 30*, 277–290.

Kahana, M.J. (1996). Associate retrieval processes in free recall. *Memory & Cognition, 24*, 103–109.

Kalbfleisch, P.J. (1992). Deceit, distrust and the social milieu: Application of deception research in a troubled world. *Journal of Applied Communication Research, 20*, 308–334.

Kassin, S.M. (2005). On the psychology of confessions: Does innocence put innocents at risk? *American Psychologist, 60*, 215–228.

Kassin, S.M. (2008a). Confession evidence: Commonsense myths and misconceptions. *Criminal Justice and Behavior, 35*, 1309–1322.

Kassin, S.M. (2008b). The psychology of confessions. *Annual Review of Law and Social Sciences, 4*, 193–217.

Kassin, S.M., Appleby, S.C., & Torkildson-Perillo, J. (2010). Interviewing suspects: Practice, science, and future directions. *Legal and Criminological Psychology, 15*, 39–56.

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.

Kassin, S.M., Goldstein, C.J., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. *Law and Human Behavior, 27*, 187–203.

Kassin, S.M., & Gudjonsson, G.H. (2004). The psychology of confessions: A review of the literature and issues. *Psychological Science in the Public Interest, 5*, 33–67.

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.

Kassin, S.M., & Norwick, R.J. (2004). Why people waive their Miranda rights: The power of innocence. *Law and Human Behavior, 28*, 211–221.

Kaufmann, G., Drevland, G.C., Wessel, E., Overskeid, G., & Magnussen, S. (2003). The importance of being earnest: Displayed emotions and witness credibility. *Applied Cognitive Psychology, 17*, 21–34.

Kendon, A. (1994). Do gestures communicate? A review. *Research on Language and Social Interaction, 27*, 175–200.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge, England: Cambridge University Press.

Kleinke, C.L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin, 100*, 78–100.

Kleinmuntz, B., & Szucko, J.J. (1984). Lie detection in ancient and modern times: A call for contemporary scientific study. *American Psychologist, 39*, 766–776.

Köhnken, G. (1987). Training police officers to detect deceptive eyewitness statements. Does it work? *Social Behaviour, 2*, 1–17.

Köhnken, G. (2002). A German perspective on children's testimony. In H.L. Westcott, G.M. Davies, & R.H.C. Bull (Eds.), *Children's testimony: A handbook of psychological research and forensic practice* (pp. 233–244). Chichester, England: Wiley.

Köhnken, G. (2004). Statement validity analysis and the "detection of the truth." In P.A. Granhag, & L.A. Strömwall (Eds.), *Deception detection in forensic contexts* (pp. 41–63). Cambridge, England: Cambridge University Press.

Kowalski, R.M., Walker, S., Wilkinson, R., Queen, A., & Sharpe, B. (2003). Lying, cheating, complaining, and other aversive interpersonal behaviors: A narrative examination of the darker side

of relationships. *Journal of Social and Personal Relationships, 20,* 471–490.

Krauss, R.M. (1981). Impression formation, impression management, and nonverbal behaviors. In E.T. Higgins, C.P. Herman, & M.P. Zanna (Eds.), *Social cognition: The Ontario Symposium* (Vol. 1, pp. 323–341). Hillsdale, NJ: Erlbaum.

LaFrance, M., & Mayo, C. (1976). Racial differences in gaze behavior during conversations: Two systematic observational studies. *Journal of Personality and Social Psychology, 33,* 547–552.

LaFrance, M., & Mayo, C. (1978). Cultural aspects of nonverbal communication. *International Journal of Intercultural Relations, 2,* 71–89.

Landry, K., & Brigham, J.C. (1992). The effect of training in criteria-based content analysis on the ability to detect deception in adults. *Law and Human Behavior, 16,* 663–675.

Lane, J.D., & DePaulo, B.M. (1999). Completing Coyne's cycle: Dysphorics' ability to detect deception. *Journal of Research in Personality, 33,* 311–329.

Leal, S., Vrij, A., Mann, S., & Fisher, R. (2010). Detecting true and false opinions: The Devil's Advocate approach as a lie detection aid. *Acta Psychologica, 134,* 323–329.

Leary, M.R., & Kowalski, R.M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107,* 34–47.

Leins, D., Fisher, R.P., Vrij, A., Leal, S., & Mann, S. (in press). Using sketch-drawing to induce inconsistency: An active approach to detecting deception. *Legal and Criminological Psychology.*

Leo, R.A. (1996). Inside the interrogation room. *Journal of Criminal Law and Criminology, 86,* 266–303.

Lerner, M.J. (1980). *The belief in a just world.* New York: Plenum Press.

Levine, T.R., Asada, K.J.K., & Park, H.S. (2006). The lying chicken and the gaze avoidant egg: Eye contact, deception and causal order. *Southern Journal of Communication, 4,* 401–411.

Levine, T.R., Feeley, T.H., McCornack, S.A., Hughes, M., & Harms, C.M. (2005). Testing the effects of nonverbal behavior training on accuracy in deception detection with the inclusion of a bogus training control group. *Western Journal of Communication, 69,* 203–217.

Levine, T.R., & McCornack, S.A. (1992). Linking love and lies: A formal test of the McCornack and Parks model of deception detection. *Journal of Social and Personal Relationships, 9,* 143–154.

Levine, T.R., & McCornack, S.A. (1996a). A critical analysis of the behavioral adaptation explanation of the probing effect. *Human Communication Research, 22,* 575–588.

Levine, T.R., & McCornack, S.A. (1996b). Can behavioral adaptation explain the probing effect? *Human Communication Research, 22,* 604–613.

Levine, T.R., & McCornack, S.A. (2001). Behavioural adaptation, confidence, and heuristic-based explanations of the probing effect. *Human Communication Research, 27,* 471–502.

Levine, T.R., Park, H.S., & McCornack, S.A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect." *Communication Monographs, 66,* 125–144.

Lindholm, T. (2008). Who can judge the accuracy of eyewitness statements? A comparison of professionals and lay-persons. *Applied Cognitive Psychology, 22,* 1301–1314.

Littmann, E., & Szewczyk, H. (1983). Zu einigen Kriterien und Ergebnissen forensisch-psychologischer Glaubwürdigkeitsbegutachtung von sexuell mißbrauchten Kindern und Jugendlichen [Criteria to be examined in forensic-psychological credibility assessment in child sexual abuse cases]. *Forensia, 4,* 55–72.

Liu, M, Granhag, P.A., Landström, S., Roos af Hjelmsäter, E., Strömwall, L.A., & Vrij, A. (2010). "Can you remember what was in your pocket when you were stung by a bee?"—Eliciting cues to deception by asking the unanticipated. *The Open Criminology Journal, 3,* 31–36.

Loftus, E.F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning and Memory, 12,* 361–366.

Loftus, E.F., & Palmer, J.C. (1974). Reconstructions of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13,* 585–589.

Lord, C.G., Ross, L., & Lepper, M.R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37,* 2098–2109.

Macrae, C.N., & Bodenhausen, G.V. (2001). Social cognition: Categorical person perception. *British Journal of Psychology, 92,* 239–256.

Malone, B.E., & DePaulo, B.M. (2001). Measuring sensitivity to deception. In J.A. Hall, & F.J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 103–124). Mahwah, NJ: Erlbaum.

Mann, S., Vrij, A., & Bull, R. (2002). Suspects, lies and videotape: An analysis of authentic high-stakes liars. *Law and Human Behavior, 26,* 365–376.

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect deceit. *Journal of Applied Psychology, 89,* 137–149.

Masip, J., Sporer, S., 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.

Matarazzo, J.D., Wiens, A.N., Jackson, R.H., & Manaugh, T.S. (1970). Interviewee speech behavior under conditions of endogenously-present and exogenously-induced motivational states. *Journal of Clinical Psychology, 26,* 17–24.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92,* 350–371.

McNeill, D. (1992). *Hand and mind.* Chicago: The University of Chicago Press.

Meissner, C.A., & Kassin, S.M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and Human Behavior, 26,* 469–480.

Millar, M.G., & Millar, K.U. (1998). The effects of suspicion on the recall of cues used to make veracity judgments. *Communication Reports, 11,* 57–64.

Miller, G.R., & Stiff, J.B. (1993). *Deceptive communication.* Newbury Park, CA: Sage.

Moston, S.J. (1992). Truth or lies. *Policing, 8,* 26–39.

Moston, S.J., & Engelberg, T. (1993). Police questioning techniques in tape recorded interviews with criminal suspects. *Policing and Society, 6,* 61–75.

Moston, S.J., Stephenson, G.M., & Williamson, T.M. (1992). The effects of case characteristics on suspect behaviour during police questioning. *British Journal of Criminology, 32*, 23–39.

National Research Council. (2003). *The polygraph and lie detection*. Washington, DC: National Academic Press.

Noller, P. (1985). Video primacy—A further look. *Journal of Nonverbal Behavior, 9*, 28–47.

Ofshe, R.J., & Leo, R.A. (1997). The decision to confess falsely: Rational choice and irrational action. *Denver University Law Review, 74*, 979–1112.

O'Sullivan, M. (2007). Unicorns or Tiger Woods: Are lie detection experts myths or rarities? A response to "On Lie Detection 'Wizards'" by Bond and Uysal. *Law and Human Behavior, 31*, 117–123.

O'Sullivan, M. (2008). Home runs and humbugs: Comment on Bond and DePaulo (2008). *Psychological Bulletin, 134*, 493–497.

O'Sullivan, M., & Ekman, P. (2004). The wizards of deception detection. In P.A. Granhag & L.A. Strömwall (Eds.), *Deception detection in forensic contexts* (pp. 269–286). Cambridge, England: Cambridge University Press.

O'Sullivan, M., Ekman, P., & Friesen, W.V. (1988). The effect of comparisons on detecting deceit. *Journal of Nonverbal Behavior, 12*, 203–216.

O'Sullivan, M., Frank, M.G., Hurley, C.M., & Tiwana, J. (in press). Police lie detection accuracy: The effect of lie scenario. *Law and Human Behavior*.

Park, H.S., Levine, T.R., McCornack, S.A., Morrisson, K., & Ferrara, M. (2002). How people really detect lies. *Communication Monographs, 69*, 144–157.

Patterson, M.L. (1995). Invited article: A parallel process model of nonverbal communication. *Journal of Nonverbal Behavior, 19*, 3–29.

Patterson, M.L. (2006). The evolution of theories of interactive behavior. In V. Manusov, & M.L. Patterson (Eds.), *The SAGE handbook of nonverbal communication* (pp. 21–39). Thousand Oaks, CA: Sage.

Pickel, K.L. (2004). When a lie becomes the truth: The effects of self-generated misinformation on eyewitness memory. *Memory, 12*, 14–26.

Porter, S., England, L., Juodis, M., ten Brinke, L., & Wilson, K. (2008). Is the face the window to the soul?: Investigation of the accuracy of intuitive judgments of the trustworthiness of human faces. *Canadian Journal of Behavioural Science, 40*, 171–177.

Porter, S., Gustaw, C., & ten Brinke, L. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology Crime & Law, 16*, 477–491.

Porter, S., Juodis, M., ten Brinke, L., Klein, R., & Wilson, K. (2010). Evaluation of a brief deception detection training program. *Journal of Forensic Psychiatry & Psychology, 21*, 66–76.

Porter, S., & Peace, K. (2007). The scars of memory: A prospective, longitudinal investigation of the consistency of traumatic and positive emotional memories in adulthood. *Psychological Science, 18*, 435–441.

Porter, S., Spencer, L., & Birt, A. (2003). Blinded by emotion? Effect of emotionality of a scene on susceptibility to false memories. *Canadian Journal of Behavioral Science, 35*, 165–175.

Porter, S., & ten Brinke, L. (2008). Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological Science, 19*, 508–514.

Porter, S., & ten Brinke, L. (2009). Dangerous decisions: A theoretical framework for understanding how judges assess credibility in the courtroom. *Legal and Criminological Psychology, 14*, 119–134.

Porter, S., & ten Brinke, L. (2010). Truth about lies: What works in detecting high-stakes deception? [Special issue]. *Legal and Criminological Psychology, 15*, 57–76.

Porter, S., & Woodworth, M. (2007). I'm sorry I did it...but he started it: A comparison of the official and self-reported homicide descriptions of psychopath and non-psychopaths. *Law and Human Behavior, 31*, 91–107.

Porter, S., Woodworth, M., & Birt, A.R. (2000). Truth, lies, and videotape: An investigation of the ability of federal parole officers to detect deception. *Law and Human Behavior, 24*, 643–658.

Porter, S., Woodworth, M., McCabe, S., & Peace, K.A. (2007). "Genius is 1% inspiration and 99% perspiration"...or is it? An investigation of the impact of motivation and feedback on deception detection. *Legal and Criminological Psychology, 12*, 297–310.

Porter, S., & Yuille, J.C. (1995). Credibility assessment of criminal suspects through statement analysis. *Psychology, Crime and Law, 1*, 319–331.

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.

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.

Quas, J.A., Davis, E.L., Goodman, G.S., & Myers, J.E.B. (2007). Repeated questions, deception, and children's true and false reports of body touch. *Child Maltreatment, 12*, 60–67.

R. v. Lifchus, 3 S.C.R. 320 (1997).

Reinhard, M.A., & Sporer, S.L. (2008). Verbal and nonverbal behaviour as a basis for credibility attribution: The impact of task involvement and cognitive capacity. *Journal of Experimental Social Psychology, 44*, 477–488.

Reinhard, M.A., Sporer, S.L., & Marksteiner, T. (2009). *Situational familiarity and the use of nonverbal and verbal information in judgments of veracity*. Manuscript under review.

Riggio, R.E. (1994). Epilogue: Where are we going, and how do we get there? *Journal of Language and Social Psychology, 13*, 514–518.

Riggio, R.E., Tucker, J., & Throckmorton, B. (1988). Social skills and deception ability. *Personality and Social Psychology Bulletin, 13*, 568–577.

Robinson, W.P. (1994). Reactions to falsifications in public and interpersonal contexts. *Journal of Language and Social Psychology, 13*, 497–513.

Roney, J.R., Hanson, K.N., Durante, K.M., & Maestripieri, D. (2006). Reading men's faces: Women's mate attractiveness judgments track men's testosterone and interest in infants. *Proceedings of the Royal Society of London B, 273*, 2169–2175.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances*

in experimental psychology (Vol. 10, pp. 174–221). New York: Academic Press.

Rotenberg, K.J., & Sullivan, C. (2003). Children's use of gaze and limb movement cues to infer deception. *The Journal of Genetic Psychology, 164*, 175–187.

Ruby, C.L., & Brigham, J.C. (1997). The usefulness of the criteria-based content analysis technique in distinguishing between truthful and fabricated allegations. *Psychology, Public Policy, and Law, 3*, 705–737.

Schacter, D.L., & Addis, R.D. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society, 392*, 773–786.

Schlenker, B.R., & Leary, M.R. (1982). Social anxiety and self-presentation: A conceptualization and model. *Psychological Bulletin, 92*, 641–669.

Schubert, S. (2006). A look tells all. *Scientific American Mind*. Retrieved June 20, 2007, from http://www.scientificamerican.com/article.cfm?id=a-look-tells-all

Schweitzer, M.E., Brodt, S.E., & Croson, R.T.A. (2002). Seeing and believing: Visual access and the strategic use of deception. *The International Journal of Conflict Management, 13*, 258–275.

Schweitzer, M.E., Hershey, J.C., & Bradlow, E.T. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes, 101*, 1–9.

Seiter, J.S., Bruschke, J., & Bai, C. (2002). The acceptability of deception as a function of perceivers' culture, deceiver's intention, and deceiver–deceived relationship. *Western Journal of Communication, 66*, 158–180.

Spence, S.A. (2008). Playing devil's advocate: The case against fMRI lie detection. *Legal and Criminological Psychology, 13*, 11–26.

Spence, S.A., Farrow, T.F.D., Herford, A.E., Wilkinson, I.D., Zheng, Y., & Woodruff, P.W.R. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport: For Rapid Communication of Neuroscience Research, 12*, 2849–2853.

Sporer, S.L. (2004). Reality monitoring and detection of deception. In P.A. Granhag, & L.A. Strömwall (Eds.), *Deception detection in forensic contexts* (pp. 64–102). Cambridge, England: Cambridge University Press.

Sporer, S.L., Penrod, S.D., Read, J.D., & Cutler, B.L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118*, 315–327.

Sporer, S.L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology, 20*, 421–446.

Sporer, S.L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law, 13*, 1–34.

Stiff, J.B., Corman, S., Krizek, B., & Snider, E. (1994). Individual differences and changes in nonverbal behavior: Unmasking the changing faces of deception. *Communication Research, 21*, 555–581.

Stiff, J.B., Kim, H.J., & Ramesh, C.N. (1992). Truth biases and aroused suspicion in relational deception. *Communication Research, 19*, 326–345.

Stiff, J.B., & Miller, G.R. (1986). "Come to think of it...": Interrogative probes, deceptive communication, and deception detection. *Human Communication Research, 12*, 339–357.

Stiff, J.B., Miller, G.R., Sleight, C., Mongeau, P., Garlick, R., & Rogan, R. (1989). Explanations for visual cue primacy in judgments of honesty and deceit. *Journal of Personality and Social Psychology, 56*, 555–564.

Stix, G. (2008, August 13). Can fMRI really tell if you are lying? *Scientific American*. Retrieved September 23, 2010, from http://www.scientificamerican.com/article.cfm?id=new-lie-detector

Strömwall, L.A., & Granhag, P.A. (2005). Children's repeated lies and truths: Effects on adults' judgments and reality monitoring scores. *Psychiatry, Psychology, & Law, 12*, 345–356.

Strömwall, L.A., & Granhag, P.A. (2007). Detecting deceit in pairs of children. *Journal of Applied Social Psychology, 37*, 1285–1304.

Strömwall, L.A., Granhag, P.A., & Hartwig, M. (2004). Practitioners' beliefs about deception. In P.A. Granhag, & L.A. Strömwall (Eds.), *Deception detection in forensic contexts* (pp. 229–250). Cambridge, England: Cambridge University Press.

Strömwall, L.A., Granhag, P.A., & Jonsson, A.C. (2003). Deception among pairs: "Let's say we had lunch together and hope they will swallow it." *Psychology, Crime & Law, 9*, 109–124.

Strömwall, L.A, Granhag, P.A., & Landström, S. (2007). Children's prepared and unprepared lies: Can adults see through their strategies? *Applied Cognitive Psychology, 21*, 457–471.

Taylor, R., & Hick, R.F. (2007). Believed cues to deception: Judgments in self-generated serious and trivial situations. *Legal and Criminological Psychology, 12*, 321–332.

Tesser, A. (1978). Self-generated attitude change. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 288–338). New York: Academic Press.

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological Inquiry, 1*, 285–293.

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences, 1124*, 208–224.

Toris, C., & DePaulo, B.M. (1984). Effects of actual deception and suspiciousness of deception on interpersonal perceptions. *Journal of Personality and Social Psychology, 47*, 1063–1073.

Trankell, A. (1972). *Reliability of evidence*. Stockholm, Sweden: Beckmans.

Trovillo, P.V. (1939). A history of lie detection. *Journal of Criminal Law and Criminology, 29*, 848–881.

Turner, R.E., Edgley, C., & Olmstead, G. (1975). Information control in conversations: Honesty is not always the best policy. *Kansas Journal of Sociology, 11*, 69–89.

Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Aussagen [Evaluation of the credibility of statements]. In U. Undeutsch (Ed.), *Handbuch der Psychologie Vol. 11: Forensische Psychologie* [Handbook of Psychology Vol. 11: Forensic psychology] (pp. 26–181). Göttingen, Germany: Hogrefe.

Undeutsch, U. (1982). Statement reality analysis. In A. Trankell (Ed.), *Reconstructing the past: The role of psychologists in criminal trials* (pp. 27–56). Deventer, The Netherlands: Kluwer.

Undeutsch, U. (1984). Courtroom evaluation of eyewitness testimony. *International Review of Applied Psychology, 33*, 51–67.

Undeutsch, U. (1989). The development of statement reality analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 101–121). Dordrecht, The Netherlands: Kluwer.

Van Rossum, W. (1998). *Verschijnen voor de rechter: Hoe het hoort en het ritueel van Turkse verdachten in de rechtszaal* [Appearing before the court: How to behave and the ritual of Turkish defendants in the courtroom]. Amsterdam, The Netherlands: Uitgeverij Duizend & Een.

Vrij, A. (1993). Credibility judgments of detectives: The impact of nonverbal behavior, social skills and physical characteristics on impression formation. *Journal of Social Psychology, 133*, 601–611.

Vrij, A. (1994). The impact of information and setting on detection of deception by police detectives. *Journal of Nonverbal Behavior, 18*, 117–137.

Vrij, A. (1995). Behavioral correlates of deception in a simulated police interview. *Journal of Psychology: Interdisciplinary and Applied, 129*, 15–29.

Vrij, A. (2000). Telling and detecting lies as a function of raising the stakes. In C.M. Breur, M.M. Kommer, J.F. Nijboer, & J.M. Reintjes (Eds.), *New trends in criminal investigation and evidence* (Vol. 2, pp. 699–709). Antwerpen, Belgium: Intersentia.

Vrij, A. (2004). Invited article: Why professionals fail to catch liars and how they can improve. *Legal and Criminological Psychology, 9*, 159–181.

Vrij, A. (2005). Criteria-based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law, 11*, 3–41.

Vrij, A. (2006). Challenging interviewees during interviews: The potential effects on lie detection. *Psychology, Crime & Law, 12*, 193–206.

Vrij, A. (2007). Deception: A social lubricant and a selfish act. In K. Fiedler (Ed.), *Frontiers of social psychology: Social communication* (pp. 309–342). New York: Psychology Press.

Vrij, A. (2008a). *Detecting lies and deceit: Pitfalls and opportunities*. (2nd ed.). Chichester, England: Wiley.

Vrij, A. (2008b). Nonverbal dominance versus verbal accuracy in lie detection: A plea to change police practice. *Criminal Justice and Behavior, 35*, 1323–1336.

Vrij, A., Akehurst, L., & Knight, S. (2006). Police officers', social workers', teachers' and the general public's beliefs about deception in children, adolescents and adults. *Legal and Criminological Psychology, 11*, 297–312.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and Human Behaviour, 26*, 261–283.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004a). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research, 30*, 8–41.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004b). Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching and deception. *Canadian Journal of Behavioural Science, 36*, 113–126.

Vrij, A., Dragt, A.W., & Koppelaar, L. (1992). Interviews with ethnic interviewees: Nonverbal communication errors in impression formation. *Journal of Community and Applied Social Psychology, 2*, 199–209.

Vrij, A., Edward, K., & Bull, R. (2001). Stereotypical verbal and nonverbal responses while deceiving others. *Personality and Social Psychology Bulletin, 27*, 899–909.

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.

Vrij, A., Evans, H., Akehurst, L., & Mann, S. (2004). Rapid judgements in assessing verbal and nonverbal cues: Their potential for deception researchers and lie detection. *Applied Cognitive Psychology, 18*, 283–296.

Vrij, A., & Fischer, A. (1995). The expression of emotions in simulated rape interviews. *Journal of Police and Criminal Psychology, 10*, 64–67.

Vrij, A., & Fischer, A. (1997). The role of displays of emotions and ethnicity in judgements of rape victims. *International Review of Victimology, 4*, 255–265.

Vrij, A., & Graham, S. (1997). Individual differences between liars and the ability to detect lies. *Expert Evidence, 5*, 144–148.

Vrij, A., & Granhag, P.A. (2007). Interviewing to detect deception. In S.A. Christianson (Ed.), *Offenders' memories of violent crimes* (pp. 279–304). Chichester, England: Wiley.

Vrij, A., Granhag, P.A., & Mann, S. (in press). Invited article: Good liars. *Journal of Psychiatry and Law*.

Vrij, A., Granhag, P.A., Mann, S., & Leal, S. (in press). Lying about flying: The first experiment to detect false intent. *Psychology, Crime & Law*.

Vrij, A., Harden, F., Terry, J., Edward, K., & Bull, R. (2001). The influence of personal characteristics, stakes and lie complexity on the accuracy and confidence to detect deceit. In R. Roesch, R.R. Corrado, & R.J. Dempster (Eds.), *Psychology in the courts: International advances in knowledge* (pp. 289–304). London: Routledge.

Vrij, A., Kneller, W., & Mann, S. (2000). The effect of informing liars about criteria-based content analysis on their ability to deceive CBCA-raters. *Legal and Criminological Psychology, 5*, 57–70.

Vrij, A., Leal, S., Granhag, P.A., Mann, S., Fisher, R.P., Hillman, J., et al. (2009). Outsmarting the liars: The benefit of asking unanticipated questions. *Law and Human Behavior, 33*, 159–166.

Vrij, A., Leal, S., Mann, S., & Granhag, P.A. (in press). A comparison between lying about intentions and past activities: Verbal cues and detection accuracy. *Applied Cognitive Psychology*.

Vrij, A., Leal, S., Mann, S., Warmelink, L., Granhag, P.A., & Fisher, R.P. (2010). Drawings as an innovative and successful lie detection tool. *Applied Cognitive Psychology, 4*, 587–594.

Vrij, A., & Mann, S. (2001a). Telling and detecting lies in a high-stake situation: The case of a convicted murderer. *Applied Cognitive Psychology, 15*, 187–203.

Vrij, A., & Mann, S. (2001b). Who killed my relative? Police officers' ability to detect real-life high-stake lies. *Psychology, Crime & Law, 7*, 119–132.

Vrij, A., & Mann, S. (2004). Detecting deception: The benefit of looking at a combination of behavioral, auditory and speech content

related cues in a systematic manner. *Group Decision and Negotiation, 13*, 61–79.

Vrij, A., Mann, S., & Fisher, R. (2006a). An empirical test of the Behaviour Analysis Interview. *Law and Human Behavior, 30*, 329–345.

Vrij, A., Mann, S., & Fisher, R. (2006b). Information-gathering vs accusatory interview style: Individual differences in respondents' experiences. *Personality and Individual Differences, 41*, 589–599.

Vrij, A., Mann, S., Fisher, R., Leal, S., Milne, B., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior, 32*, 253–265.

Vrij, A., Mann, S., Kristen, S., & Fisher, R. (2007). Cues to deception and ability to detect lies as a function of police interview styles. *Law and Human Behavior, 31*, 499–518.

Vrij, A., Mann, S., Leal, S., & Fisher, R. (2010). "Look into my eyes": Can an instruction to maintain eye contact facilitate lie detection? *Psychology, Crime & Law, 16*, 327–348.

Vrij, A., Mann, S., Leal, S., & Granhag, P.A. (2010). Getting into the minds of pairs of liars and truth tellers: An examination of their strategies [Special issue]. *The Open Criminology Journal, 3*, 17–22.

Vrij, A., & Semin, G.R. (1996). Lie experts' beliefs about nonverbal indicators of deception. *Journal of Nonverbal Behavior, 20*, 65–80.

Vrij, A., & Winkel, F.W. (1991). Cultural patterns in Dutch and Surinam nonverbal behavior: An analysis of simulated police/citizen encounters. *Journal of Nonverbal Behavior, 15*, 169–184.

Vrij, A., & Winkel, F.W. (1992). Crosscultural police–citizen interactions: The influence of race, beliefs and nonverbal communication on impression formation. *Journal of Applied Social Psychology, 22*, 1546–1559.

Vrij, A., & Winkel, F.W. (1994). Perceptual distortions in cross-cultural interrogations: The impact of skin color, accent, speech style and spoken fluency on impression formation. *Journal of Cross-Cultural Psychology, 25*, 284–296.

Vrij, A., Winkel, F.W., & Koppelaar, L. (1991). Interactie tussen politiefunctionarissen en allochtone burgers: twee studies naar de frequentie en het effect van aan- en wegkijken op de impressieformatie [Cross-cultural police–citizen interactions: Two experiments examining the impact of gaze behavior on impression formation]. *Nederlands Tijdschrift voor de Psychologie, 46*, 8–20.

Waenke, M., & Bless, H. (2000). The effects of subjective ease of retrieval on attitudinal judgements: The moderating role of processing motivation. In H. Bless, & J.P. Forgas (Eds.), *The message within: The role of subjective experience in social cognition and behaviour* (pp. 143–161). Philadelphia, PA: Psychology Press.

Walczyk, J.J., Roper, K.S., Seemann, E., & Humphrey, A.M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology, 17*, 755–774.

Walczyk, J.J., Schwartz, J.P., Clifton, R., Adams, B., Wei, M., & Zha, P. (2005). Lying person-to-person about live events: A cognitive framework for lie detection. *Personnel Psychology, 58*, 141–170.

Wang, G., Chen, H., & Atabakhsh, H. (2004). Criminal identity deception and deception detection in law enforcement. *Group Decision and Negotiation, 13*, 111–127.

Watanabe, H. (2005). Semantic and episodic predictions of memory for plans. *Japanese Psychological Research, 47*, 40–45.

Wessel, E., Drevland, G., Eilertsen, D.E., & Magnussen, S. (2006). Credibility of the emotional witness: A comparison of ratings by laypersons and court judges. *Law and Human Behavior, 30*, 221–230.

White, C.H., & Burgoon, J.K. (2001). Adaptation and communicative design: Patterns of interaction in truthful and deceptive conversations. *Human Communication Research, 27*, 9–37.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*, 592–598.

Winkel, F.W., & Koppelaar, L. (1991). Rape victims' style of self-presentation and secondary victimization by the environment. *Journal of Interpersonal Violence, 6*, 29–40.

Winkel, F.W., & Vrij, A. (1990). Interaction and impression formation in a cross-cultural dyad: Frequency and meaning of culturally determined gaze behaviour in a police interview setting. *Social Behaviour, 5*, 335–350.

Wolpe, P.R., Foster, K.R., & Langleben, D.D. (2005). Emerging neurotechnologies for lie-detection: Promises and perils. *The American Journal of Bioethics, 5*, 39–49.

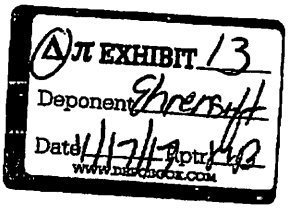
Yuille, J.C. (1988). The systematic assessment of children's testimony. *Canadian Psychology, 29*, 247–262.

Zuckerman, M., DePaulo, B.M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1–57). New York: Academic Press.

Zuckerman, M., Driver, R., & Koestner, R. (1982). Discrepancy as a cue to actual and perceived deception. *Journal of Nonverbal Behavior, 7*, 95–100.

Zuckerman, M., Speigel, N.H., DePaulo, B.M., & Rosenthal, R. (1982). Nonverbal strategies for decoding deception. *Journal of Nonverbal Behavior, 6*, 171–187.

Court's Exhibit 5M
(Under Seal)



PSYCHOLOGICAL SCIENCE IN THE PUBLIC INTEREST

Editorial

Connecting Clinical Practice to Scientific Progress

Walter Mischel

Columbia University

Paul Meehl, in one of his last public speeches, memorably noted that most clinical psychologists select their methods like kids make choices in a candy store: They look around, maybe sample a bit, and choose what they like, whatever feels good to them. For many of us who initially became clinical psychologists because we were inspired by the scientist-practitioner ideal, Meehl's comment was as heartbreaking as it was accurate. It makes particularly compelling the article that follows, "Current Status and Future Prospects of Clinical Psychology: Toward a Scientifically Principled Approach to Mental and Behavioral Health Care" by Baker, McFall, and Shoham. This urgently needed and long overdue analysis and proposal will be welcomed by those who grieve the widening gulf between clinical practice and scientific progress in psychology. And it offers giant but feasible steps toward reforms that can advance both clinical practice and relevant psychological science, to at last reverse the disconnect that has been unfortunate for each.

The authors' proposal for a "scientifically principled approach to mental and behavioral health care" is an incisive and scholarly analysis of where clinical psychology is (and is not) today, how it got there, and how it will increasingly discredit and marginalize itself if it continues the trajectory it has pursued for far too many years. But it is also much more. The article makes clear the heavy costs and consequences to the profession, and more important to the people who have a right to expect much more from their health care providers. Most exciting, it charts a route toward a scientifically principled and thus responsible approach to the mental and behavioral health care that our science can offer and that those who suffer from mental and behavioral problems deserve to get.

The disconnect between much of clinical practice and the advances in psychological science is an unconscionable embarrassment for many reasons, and a case of professional cognitive dissonance with heavy costs. The Boulder Model of the scientist-practitioner, now mostly a historical footnote and a cue for depression, came half a century ago when psychological science was still somewhere between its infancy and its turbulent adolescence. Evidence for most assessment and treatment methods for clinical psychology was still far from solid, and usually highly dubious, making the choices of practitioners "like kids in a candy store" more understandable. The distressing cognitive dissonance now is that the science has advanced dramatically over the last 50 years, and there are now numerous state-of-the-science–based and empirically supported choices for assessment and for treatment, yet practitioners too often still choose to do whatever they feel like, as Meehl described, regardless of evidence.

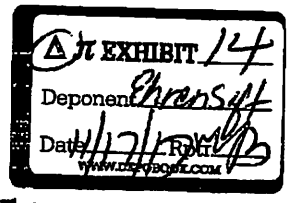
In my own career, I struggled with these issues beginning in the 1960s. During many of my 20 years at Stanford University, Albert Bandura and I tried to hold on to a science-based clinical training program. The bizarre situation we faced there is of more than personal and historical interest: I suspect that many of the same conflicts still exist and motivate the efforts described by Baker and colleagues. Bandura and I, and our students and other colleagues, were discovering the remarkable discrepancies between what the scientific work was revealing and the requirements imposed by the pressures for maintaining accreditation. The professional accreditation requirements insisted on continuing practices whose value was contradicted by the empirical findings. Those requirements not only flew in the face of the data but also made enormous demands on faculty and student time in the clinical program. At one point, Bandura made a table of faculty arrivals and departures in our clinical program. It showed rapid, continuous turnover among the junior faculty in clinical, because those who devoted their time to clinical work and were good at it generally did not meet the academic standards, and vice versa, so accepting a clinical position at Stanford almost guaranteed no future in the university. For a temporary solution, we turned the clinical program into one on experimental psychopathology. It included more experimental work and research, most of it within clinical settings and directly relevant to clinical applications. In it we also could move away from techniques that neither of us believed in, given the data, and that both of us were trying to change—from costly tests with little or no validity to therapies without evidence of efficacy but on which the American Psychological Association insisted for clinical programs and for acceptable internship experiences. It became a program that helped train many of the people who

became leaders in the development of cognitive behavior modification and assessment. And as the pressures grew, it became impossible to maintain.

Baker, McFall, and Shoham make a compelling case for what many of us have long believed: A realistic route for change requires a new accreditation system that demands high-quality science training and insists on it as part of the core for doctoral training in clinical psychology. The good news—the first in a very long time on this topic—is that such a system is here in the new Psychological Clinical Science Accreditation System (PCSAS). Its mission is to "accredit clinical psychology training programs that offer high quality science-centered education and training, producing graduates who are successful in generating and applying scientific knowledge" (p. ii). It is a mission that deserves the strongest support.

Support for the movement toward a scientifically principled clinical psychology has self-evident potential benefits to the public, to the profession, and to our science. It's also worth remembering that many of our best students still enter psychology to become clinical psychologists. They deserve the opportunity to do such work informed and guided by evidence, trained to evaluate it properly, and able to add to it themselves.

PSYCHOLOGICAL SCIENCE



Research Report

Psychophysiological Responding During Script-Driven Imagery in People Reporting Abduction by Space Aliens

Richard J. McNally,[1] Natasha B. Lasko,[2,3,4] Susan A. Clancy,[1] Michael L. Macklin,[2] Roger K. Pitman,[3,4] and Scott P. Orr[2,3,4]

[1]Harvard University; [2]Veterans Affairs Medical Center, Manchester, New Hampshire; [3]Massachusetts General Hospital; and [4]Harvard Medical School

ABSTRACT—*Is recollection of highly improbable traumatic experiences accompanied by psychophysiological responses indicative of intense emotion? To investigate this issue, we measured heart rate, skin conductance, and left lateral frontalis electromyographic responses in individuals who reported having been abducted by space aliens. Recordings of these participants were made during script-driven imagery of their reported alien encounters and of other stressful, positive, and neutral experiences they reported. We also measured the psychophysiological responses of control participants while they heard the scripts of the abductees. We predicted that if "memories" of alien abduction function like highly stressful memories, then psychophysiological reactivity to the abduction and stressful scripts would be greater than reactivity to the positive and neutral scripts, and this effect would be more pronounced among abductees than among control participants. Contrast analyses confirmed this prediction for all three physiological measures (*ps *< .05). Therefore, belief that one has been traumatized may generate emotional responses similar to those provoked by recollection of trauma (e.g., combat).*

Few controversies in psychology have been as contentious as the one concerning recovered memories of trauma (McNally, 2003b). Especially contentious has been the claim that some people may recover "false memories" of traumatic events that never occurred (e.g., Ceci & Loftus, 1994; Lindsay & Read, 1994). Only recently, however, have researchers begun to study memory function in people reporting recovered memories of trauma (e.g., Clancy, McNally, & Schacter, 1999; McNally, 2003a; McNally, Clancy, Barrett, & Parker, in press; McNally, Clancy, & Schacter, 2001). Adapting Roediger and McDermott's (1995) procedure, we found that adults reporting recovered memories of childhood sexual abuse were more likely to exhibit false recognition of nonpresented words than were adults who reported always remembering their abuse (Clancy, Schacter, McNally, & Pitman, 2000). A subsequent study revealed similar false memory effects in people reporting recovered memories of alien abduction (Clancy, McNally, Schacter, Lenzenweger, & Pitman, 2002).

People who have developed posttraumatic stress disorder (PTSD) usually exhibit heightened psychophysiological reactivity (e.g., increased heart rate, HR) when recalling their trauma in the laboratory (for a review, see Orr, Metzger, & Pitman, 2002). Clinical reports suggest that recovering memories of improbable traumatic events (e.g., being ritually abused by satanic cults) is likewise accompanied by intense emotional reactions (e.g., Young, Sachs, Braun, & Watkins, 1991), and some therapists interpret these reactions as evidence that something horrific must have happened to the person (e.g., Bloom, 1994).

In the present study, we investigated whether recollection of highly improbable traumatic events provokes psychophysiological reactions indicative of intense emotion. We recruited individuals who reported having been abducted by space aliens and asked them to participate in a script-driven imagery protocol (e.g., Lang, Levin, Miller, & Kozak, 1983; Pitman, Orr, Forgue, de Jong, & Claiborn, 1987). Each abductee furnished material for five personalized, autobiographical scripts: two scripts related to his or her abduction trauma; a script related to a different, extremely stressful experience; a script related to an extremely positive experience; and one related to an emotionally neutral experience. A control group consisted of individuals who denied ever having been abducted by aliens, but who listened to and imagined the scripts provided by the abductees. We predicted that if "memories" of alien abduction function like highly stressful

Address correspondence to Richard J. McNally, Department of Psychology, Harvard University, 33 Kirkland St., Cambridge, MA 02138; e-mail: rjm@wjh.harvard.edu.

memories, then psychophysiological reactivity to the abduction and stressful scripts would be greater than reactivity to the positive and neutral scripts, and this effect would be more pronounced among abductees than among control participants.

METHOD

Participants

Alien-Abductee Group

The alien-abductee group comprised 6 women and 4 men who reported having been abducted by alien beings. Their mean age was 47.5 years ($SD = 11.9$). They learned of our research program on the "psychophysiology of emotional memory" through newspaper advertisements; staff at the Program for Extraordinary Experience Research (PEER), Center for Psychology and Social Change, Cambridge, Massachusetts; or previous participants. Recruitment and testing were in accordance with the American Psychological Association's ethical guidelines regarding the use of human participants. The protocol and informed-consent form was approved by the Harvard University Committee on the Use of Human Subjects and by the Manchester Veterans Affairs Medical Center Human Subjects Committee.

During the first session, each participant was interviewed by either Richard J. McNally or Susan A. Clancy about his or her encounters with space aliens. The participant then completed the script-preparation forms (described later). During the second session, Natasha B. Lasko used the Clinician-Administered PTSD Scale-Diagnostic Version (CAPS; Blake et al., 1995) and the Structured Clinical Interview for Axis I DSM-IV Disorders (SCID; First, Spitzer, Gibbon, & Williams, 1994) to assess for PTSD related to purported alien abduction and to assess for other Axis I disorders.

Three participants fell short, by one or two symptoms, of qualifying for lifetime PTSD related to their alien encounters, and 1 of these individuals had current subthreshold PTSD.

All abductees reported at least one episode of apparent sleep paralysis accompanied by hypnopompic hallucinations, usually figures hovering near their beds, flashing lights, buzzing sounds, and tingling sensations. In each case, the participant interpreted the experience as related to aliens. Eight of the 10 abductees had undergone quasi-hypnotic sessions during which mental health professionals helped them recover detailed "memories" of alien encounters (e.g., undergoing sexual and medical probing on spaceships).

Control Group

The control group comprised 7 women and 5 men, recruited from the community. Their mean age was 49.9 years ($SD = 13.0$).

Psychometrics

Participants completed several questionnaires: the Dissociative Experiences Scale (DES; Bernstein & Putnam, 1986); the Beck Depression Inventory (BDI; Beck & Steer, 1987); the Trait Anxiety Inventory (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983); the Absorption Scale, a measure of imaginative capability and fantasy proneness (Tellegen & Atkinson, 1974); and measures of schizotypy: the Perceptual Aberration Scale (Chapman, Chapman, & Raulin, 1978), the Magical Ideation Scale (Eckblad & Chapman, 1983), and the Referential Thinking Scale (Lenzenweger, Bennett, & Lilenfeld, 1997).

Scripts

Following the abduction-history interview, we prepared five individualized scripts describing autobiographical events from each abductee's past: two scripts related to alien abduction, one stressful script (unrelated to abduction), one positive script (unrelated to abduction), and one neutral script (unrelated to abduction). Per standard procedure (Orr et al., 1998), we averaged the responses for the two abduction scripts prior to data analysis.

The abductees first described each event on a script-preparation form and then selected from a list of bodily responses those that they remembered experiencing when the event was occurring. Using this information, we prepared 30-s (approximately) scripts describing each experience in the second person, present tense. Each script referred to the bodily responses endorsed by the subject and incorporated words and phrases used by the subject on the script-preparation form. The scripts were audiotaped for playback in the psychophysiology laboratory.

Episodes of apparent sleep paralysis, interpreted as alien encounters, figured in both abduction trauma scripts for 3 abductees and in one abduction script for 2 additional abductees; for the remaining 5 abductees, both abduction scripts featured traumatic experiences (e.g., being sexually probed by aliens on board spaceships) that typically surfaced during quasi-hypnotic recovered-memory sessions. Examples of stressful, positive, and neutral scripts were learning of the violent death of loved ones, witnessing the birth of one's first child, and cutting one's lawn during the previous weekend, respectively.

Each participant in the control group heard the scripts of one of the abductees. Each control participant was matched with an abductee of the same sex and age. This yoking procedure controlled for materials effects (i.e., the possibility that anyone listening to scripts of alien abduction might exhibit psychophysiological reactivity).

Apparatus and Physiological Variables

The psychophysiology session was conducted in an 11- × 9-ft humidity- and temperature-controlled, sound-attenuated testing room connected via wires to an adjacent room where the apparatus was located. The participant sat in a comfortable armchair. A monitor in the subject's room displayed the self-report scales for emotion, and participants' self-reports were entered into the computer via a joystick. A modular instrument system (Coulbourn Instruments, Allentown, Pennsylvania) recorded analog physiological signals, which were monitored by V-212 oscilloscopes (Hitachi Denshi, Ltd., Tokyo, Japan).

Dependent physiological variables included HR, skin conductance (SC), and electromyogram (EMG) of the left lateral frontalis (LF) facial muscle. EMG was obtained through 4-mm (sensor diameter) silver/silver chloride electrodes filled with an electrolytic paste, placed according to standard specifications (Fridlund & Cacioppo, 1986), attached to a bioamplifier (Coulbourn Hi-Gain, S75-01), and integrated via a 300-ms time constant through a contour-following integrator (Coulbourn, S76-01). SC measurements were obtained through 9-mm (sensor diameter) silver/silver chloride electrodes filled with an isotonic paste, placed on the subject's nondominant palm, and connected to an SC module (Coulbourn, S71-11), which used a constant voltage (0.5 V) in the direct-coupled mode (Fowles et al., 1981). HR

measurements were obtained via standard limb electrocardiogram leads connected to a bioamplifier (Coulbourn Hi-Gain, S75-01) that provided input to a tachometer (Coulbourn, S77-26). Analog outputs of the physiological modules were digitized by an analog-to-digital converter (Coulbourn, S25-12) prior to sampling. A personal computer controlled presentation of the audiotaped scripts, administration of the emotion self-report scales, and sampling and storing of the digitized physiological signals at 2 Hz. A Coulbourn Lablink Computer Interface connected the computer to the instrument system.

Procedure and Data Reduction

After receiving an orientation to the laboratory and having electrodes attached, participants listened to a 3-min relaxation instruction tape prior to listening to the audiotaped scripts. Each script presentation comprised four consecutive 30-s periods: baseline, listening, imagery, and recovery. Participants were told to listen carefully to each script and imagine it as vividly as possible, as if it were actually occurring (listening period), and at the end of the script to continue imagining the experience from beginning to end (imagery period) until a tone sounded. They were instructed to cease imagery upon hearing the tone and to relax (recovery period). Upon hearing a second tone, participants provided self-reports of image vividness, three dimensions of emotional response (valence, arousal, and dominance), and seven discrete emotional responses (sadness, anger, fear, disgust, surprise, happiness, guilt). These self-reports were made on 13-point Likert scales ranging from 0 (*none*) to 12 (*a great deal*). The computer was programmed to begin the baseline period for the next script after a rest period of 1 min or when the HR of the subject had returned to within 5% of its value during the previous baseline period, whichever was longer. The rest period seldom exceeded 3 min.

The mean level of each physiological variable was computed for each data-collection period for each script. As in previous work (e.g., Pitman et al., 1987), we calculated change scores by subtracting the preceding baseline-period value from the value for the imagery period that followed it. Because of recording problems, the HR data for 1 control participant could not be used.

RESULTS

Psychometrics

Abductees scored significantly higher than control participants on measures of absorption, magical ideation, and dissociation (see Table 1).

TABLE 1
Psychometric Measures

	Group						
	Abductee		Control				
Variable	*M*	*SD*	*M*	*SD*	*t*	*df*	*p*
CAPS-L	38.2	20.4	—	—	—	—	—
CAPS-C	15.4	13.4	—	—	—	—	—
DES	8.4	7.0	3.3	3.5	2.22	19	.039
Absorption	21.6	6.0	9.6	6.1	4.50	19	.001
BDI	3.6	5.7	1.7	2.4	0.92	16	.373
Trait Anxiety	36.1	9.3	30.5	7.2	1.38	15	.189
RTS	2.9	4.1	1.6	2.0	0.87	16	.397
PAS	3.3	4.0	1.7	1.7	1.16	16	.262
MIS	9.2	4.4	2.9	2.7	3.65	16	.002

Note. CAPS-L and CAPS-C = Lifetime and Current total scores, respectively, on the Clinician-Administered PTSD Scale (Blake et al., 1995; possible range: 0–136); DES = Dissociative Experiences Scale (Bernstein & Putnam, 1986; possible range: 0–34); Absorption = Absorption Scale (Tellegen & Atkinson, 1974; possible range: 0–34); BDI = Beck Depression Inventory (Beck & Steer, 1987; possible range: 0–64); Trait Anxiety = Trait Anxiety Inventory (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983; possible range: 20–80); RTS = Referential Thinking Scale (Lenzenweger, Bennett, & Lilenfeld, 1997; possible range: 0–34); PAS = Perceptual Aberration Scale (Chapman, Chapman, & Raulin, 1978; possible range: 0–35); MIS = Magical Ideation Scale (Eckblad & Chapman, 1983; possible range: 0–30). Because of missing data, degrees of freedom vary.

Reactions to Imagery Scripts

If reported memories of alien encounters provoke reactions akin to those provoked by traumatic memories, then abductees should exhibit greater reactivity to abduction and stressful scripts than to other positive and neutral scripts, relative to control participants. To test this hypothesis, we first applied contrast weights of −1, −1, 1, and 1 to each participant's physiological response (e.g., HR increase) to the personal neutral, positive, stressful, and abduction (average of both) scripts. After multiplying each contrast weight and its respective physiological value, we created an *L* score for each participant by summing the products obtained. The larger the *L* score, the more a participant tended to produce larger responses to the abduction and stressful scripts than to the positive and neutral scripts. To test the hypothesis that this effect would be greater among abductees than among control participants, we conducted a one-tailed *t* test on the *L* scores.

The results were consistent with our hypothesis. Relative to control participants, abductees exhibited greater psychophysiological reactivity to abduction and stressful scripts than to positive and neutral scripts. This hypothesis was supported for HR, $t(19) = 2.01$, $p = .03$, effect size $r = .42$; for SC, $t(20) = 1.88$, $p = .04$, effect size $r = .39$; and for LF EMG, $t(20) = 2.00$, $p = .03$, effect size $r = .41$ (Fig. 1).[1]

Self-reported emotional responses were consistent with physiological responses in that the abductees reported heightened ratings of arousal, fear, surprise, and imagery vividness during exposure to scripts featuring their most traumatic abduction memories (see Table 2).

DISCUSSION

Recollections of purported traumatic encounters with space aliens are accompanied by physiological reactions and emotional self-reports akin to those accompanying other highly stressful experiences.[2] Relative to control participants, the abductees scored significantly higher on questionnaire measures of dissociation, absorption, and

[1] For the abductees, the abduction and stressful scripts were physiologically indistinguishable, as evinced by two-tailed paired *t* tests: HR, $t(9) = 0.57$, $p = .58$; SC, $t(9) = 0.10$, $p = .91$; LF EMG, $t(9) = 1.17$, $p = .27$. In contrast, participants with PTSD usually exhibit greater responses to trauma than to other stressful scripts (Orr & Roth, 2000).

[2] It is highly unlikely that our findings are attributable merely to the abductees' having been exposed to personalized scripts and the control participants' having been exposed to the scripts of strangers (i.e., the abductees). Indeed, personalized combat scripts are insufficient to provoke heightened physiological responses in Vietnam veterans who do not have PTSD (Orr et al., 2002).

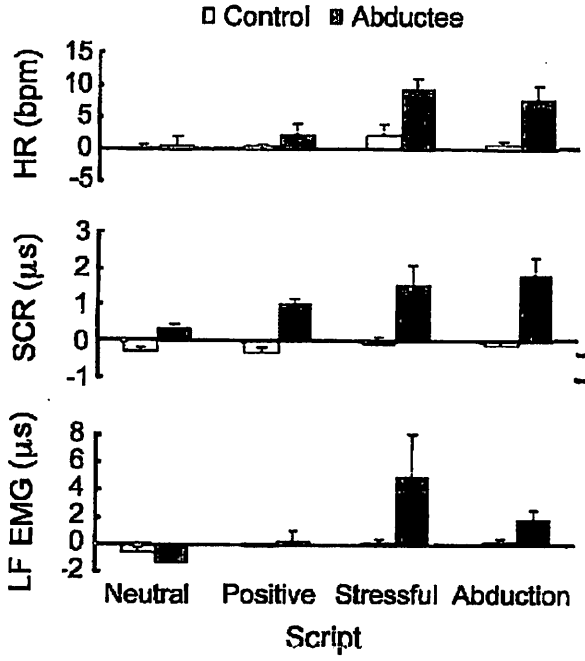


Fig. 1. Group mean (and standard error of the mean) heart rate (HR, top panel), skin conductance (SCR, middle panel), and left lateral frontalis electromyogram (LF EMG, bottom panel) responses during imagery of neutral, positive, stressful, and abduction (average of two scripts) scripts.

magical ideation. People scoring high on these three measures tend to experience alterations in consciousness, to have a rich fantasy life, and to endorse unconventional beliefs (e.g., mind reading, prophetic dreams), respectively.

TABLE 2

Self-Reported Emotional Responses During Abduction Imagery

	Group					
	Abductee		Control			
Variable	*M*	*SD*	*M*	*SD*	*t*(20)	*p*
Vividness	10.8	2.1	7.1	2.8	3.4	.003
Emotional dimensions						
Arousal	10.2	2.2	8.4	1.8	2.1	.047
Valence	1.9	2.7	3.6	3.2	1.3	.20
Dominance	2.2	2.6	4.4	4.0	1.5	.14
Discrete emotions						
Happiness	0.5	0.8	1.5	2.7	1.2	.26
Sadness	5.0	4.9	3.8	4.4	0.6	.57
Fear	9.4	3.5	5.8	4.6	2.0	.06
Surprise	9.1	3.8	5.3	3.6	2.4	.03
Anger	6.8	4.7	5.4	4.7	0.7	.48
Disgust	5.0	4.0	5.0	4.8	0.0	.98
Guilt	2.9	3.3	2.0	3.7	0.6	.54

Note. Scale ranges were as follows—arousal: 0 = *unaroused*, 12 = *aroused*; valence: 0 = *displeased*, 12 = *pleased*; dominance: 0 = *submissive*, 12 = *dominant*; vividness and discrete emotions: 0 = *none*, 12 = *a great deal*. Tests were two-tailed.

The responses of abductees to their traumatic abduction scripts bear comparison to the responses of PTSD patients to scripts of their traumatic experiences. The abductees' mean HR, SC, and LF EMG in response to their abduction scripts were 7.8 bpm, 1.8 μS, and 1.8 μV, respectively. The corresponding values for 72 PTSD participants' responses to their trauma scripts were 7.9 bpm for HR, 1.0 μS for SC, and 2.6 μV for LF EMG (Orr & Roth, 2000).

Although improbable traumatic memories (e.g., being sexually probed on a spaceship) provoke physiological reactions comparable to those provoked by more conventional and verifiable traumatic memories (e.g., a firefight in Vietnam), one should not conclude that PTSD patients are reporting false memories of trauma. Conversely, the physiological markers of emotion that accompany recollection of a memory cannot be taken as evidence of the memory's authenticity. The script-driven imagery protocol reflects the emotional significance of a memory, not necessarily its veracity.

Acknowledgments—This research was supported by a grant from the Clark Fund, Faculty of Arts and Sciences, Harvard University. We thank John E. Mack, Paul Bernstein, and the staff of the Program for Extraordinary Experience Research for publicizing our research program among potential participants. We also thank Heidi M. Barrett, Heike Croteau, Mark F. Lenzenweger, Carel S. Ristuccia, and Carol A. Perlman for their assistance.

REFERENCES

Beck, A.T., & Steer, R.A. (1987). *Beck Depression Inventory manual.* San Antonio, TX: Psychological Corp.

Bernstein, E.M., & Putnam, F.W. (1986). Development, reliability, and validity of a dissociation scale. *Journal of Nervous and Mental Disease, 174,* 727–735.

Blake, D.D., Weathers, F.W., Nagy, L.M., Kaloupek, D.G., Gusman, F.D., Charney, D.S., & Keane, T.M. (1995). The development of a clinician-administered PTSD Scale. *Journal of Traumatic Stress, 8,* 75–90.

Bloom, S.L. (1994). Hearing the survivor's voice: Sundering the wall of denial. *Journal of Psychohistory, 21,* 461–477.

Ceci, S.J., & Loftus, E.F. (1994). 'Memory work': A royal road to false memories? *Applied Cognitive Psychology, 8,* 351–364.

Chapman, L.J., Chapman, J.P., & Raulin, M.L. (1978). Body-image aberration in schizophrenia. *Journal of Abnormal Psychology, 87,* 399–407.

Clancy, S.A., McNally, R.J., & Schacter, D.L. (1999). Effects of guided imagery on memory distortion in women reporting recovered memories of childhood sexual abuse. *Journal of Traumatic Stress, 12,* 559–569.

Clancy, S.A., McNally, R.J., Schacter, D.L., Lenzenweger, M.F., & Pitman, R.K. (2002). Memory distortion in people reporting abduction by aliens. *Journal of Abnormal Psychology, 111,* 455–461.

Clancy, S.A., Schacter, D.L., McNally, R.J., & Pitman, R.K. (2000). False recognition in women reporting recovered memories of sexual abuse. *Psychological Science, 11,* 26–31.

Eckblad, M., & Chapman, L.J. (1983). Magical ideation as an indicator of schizotypy. *Journal of Abnormal Psychology, 51,* 215–225.

First, M.B., Spitzer, R.L., Gibbon, M., & Williams, J.B.W. (1994). *Structured Clinical Interview for Axis I DSM-IV Disorders* (Version 2.0). New York: New York State Psychiatric Institute, Biometrics Research Department.

Fowles, D.C., Christie, M.J., Edelberg, R., Grings, W.W., Lykken, D.T., & Venables, P.H. (1981). Publication recommendations for electrodermal measurements. *Psychophysiology, 18,* 232–239.

Fridlund, A.J., & Cacioppo, J.T. (1986). Guidelines for human electromyographic research. *Psychophysiology, 23,* 567–589.

Lang, P.J., Levin, D.N., Miller, G.A., & Kozak, M.J. (1983). Fear behavior, fear imagery, and the psychophysiology of emotion: The problem of affective-response integration. *Journal of Abnormal Psychology, 92,* 276–306.

Lenzenweger, M.F., Bennett, M.E., & Lilenfeld, L.R. (1997). The Referential Thinking Scale as a measure of schizotypy: Scale development and initial construct validation. *Psychological Assessment, 9*, 452–463.

Lindsay, D.S., & Read, J.D. (1994). Psychotherapy and memories of childhood sexual abuse: A cognitive perspective. *Applied Cognitive Psychology, 8*, 281–338.

McNally, R.J. (2003a). Recovering memories of trauma: A view from the laboratory. *Current Directions in Psychological Science, 12*, 32–35.

McNally, R.J. (2003b). *Remembering trauma*. Cambridge, MA: Belknap Press/Harvard University Press.

McNally, R.J., Clancy, S.A., Barrett, H.M., & Parker, H.A. (in press). Inhibiting retrieval of trauma cues in adults reporting histories of childhood sexual abuse. *Cognition and Emotion*.

McNally, R.J., Clancy, S.A., & Schacter, D.L. (2001). Directed forgetting of trauma cues in adults reporting repressed or recovered memories of childhood sexual abuse. *Journal of Abnormal Psychology, 110*, 151–156.

Orr, S.P., Lasko, N.B., Metzger, L.J., Berry, N.J., Ahern, C.E., & Pitman, R.K. (1998). Psychophysiologic assessment of women with posttraumatic stress disorder resulting from childhood sexual abuse. *Journal of Consulting and Clinical Psychology, 66*, 906–913.

Orr, S.P., Metzger, L.J., & Pitman, R.K. (2002). Psychophysiology of posttraumatic stress disorder. *Psychiatric Clinics of North America, 25*, 271–293.

Orr, S.P., & Roth, W.T. (2000). Psychophysiological assessment: Clinical application for PTSD. *Journal of Affective Disorders, 61*, 225–240.

Pitman, R.K., Orr, S.P., Forgue, D.F., de Jong, J.B., & Claiborn, J.M. (1987). Psychophysiologic assessment of posttraumatic stress disorder imagery in Vietnam combat veterans. *Archives of General Psychiatry, 44*, 970–975.

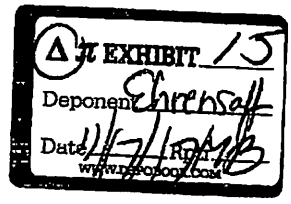
Roediger, H.L., III, & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814.

Spielberger, C.D., Gorsuch, R.L., Lushene, R., Vagg, P.R., & Jacobs, G.A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology, 83*, 268–277.

Young, W.C., Sachs, R.G., Braun, B.G., & Watkins, R.T. (1991). Patients reporting ritual abuse in childhood: A clinical syndrome. Report of 37 cases. *Child Abuse and Neglect, 15*, 181–189.

(Received 4/14/03; Revision accepted 8/27/03)



SCIENCE AND SOCIETY

Our changeable memories: legal and practical implications

Elizabeth Loftus

The malleability of memory is becoming increasingly clear. Many influences can cause memories to change or even be created anew, including our imaginations and the leading questions or different recollections of others. The knowledge that we cannot rely on our memories, however compelling they might be, leads to questions about the validity of criminal convictions that are based largely on the testimony of victims or witnesses. Our scientific understanding of memory should be used to help the legal system to navigate this minefield.

Memories are precious. They give us identity. They create a shared past that bonds us with family and friends. They seem fixed, like concrete, so that if you 'stepped' on them they would still be there as they always were.

But memories are not fixed. Everyday experience tells us that they can be lost, but they can also be drastically changed or even created. Inaccurate memories can sometimes be as compelling and 'real' as an accurate memory. In this article, I discuss the ways in which memories can be reshaped and their implications for the legal system. If we cannot believe our own memories, how can we know whether the memories of a victim or a witness are accurate?

Remaking memories

We are all familiar with temporary memory problems. "I can't remember the right word," says a colleague at a cocktail party. "Is it senility?" I reply: "Can you remember the word later?" And the usual answer will be yes, proving that the information was not lost, but only temporarily unavailable. Retrieval problems are common.

However, there are also problems with storing something new. This usually occurs simply because the person concerned is not paying attention. But some people are unable to store new information even if they are paying attention and have the opportunity to repeat the new information over and over again — several hours later, it is gone. Such people, including patients with Alzheimer's disease, might not even complain about 'losing their memory' because they do not realize that anything is missing[1].

More insidiously, memories can become scrambled, sometimes in the process of attempting to retrieve something. You might relate a story to a friend but unwittingly include some mistaken details. Later, as you attempt to recall the episode, you might come across your memory of the scrambled recall attempt instead of your original memory. Memory is malleable. It is not, as is commonly thought, like a museum piece sitting in a display case. "Memory is," as the Uruguayan novelist Eduardo Galeano once said, "born every day, springing from the past, and set against it."[2]

Usually the scrambled memory does not matter very much. But if you are an eyewitness to a crime, your scrambled recall could send someone to prison. And, rather than feeling hesitant, you might feel perfectly sure of the truth of your memory. The history of the United States justice system, like those of other countries, is littered with wrongful convictions made on the basis of mistaken memories[3]. Huff recently estimated[4] that about 7,500 people arrested for serious crimes were wrongly convicted in the United States in 1999. He further noted that the rate is thought to be much lower in Great Britain, Canada, Australia, New Zealand and many other nations, especially those that have established procedures for reviewing cases involving the potential of wrongful conviction.

Ronald Cotton, a North Carolina prisoner who was convicted in 1986 of raping a 22-year-old college student, Jennifer Thompson, puts a human face on these cases. Thompson stood up on the stand, put her hand on the Bible and swore to tell the truth. On the basis of her testimony, Cotton was sentenced to prison for life. Eventually, DNA testing — which began 11 years after Thompson had first identified Cotton — proved his innocence. Another man, Bobby Poole, pleaded guilty to the crime[3].

Faulty memory is not just about picking the wrong person. Memory problems were also evident during the sniper attacks that killed ten people in the Washington DC area in 2002 (see for example, REF. 5). Witnesses reported seeing a white truck or van fleeing several of the crime scenes. It seems that a white vehicle might have been near one of the first shootings and media repetition of this information contaminated the memories of witnesses to later attacks, making them more likely to remember white trucks. When caught, the sniper suspects were driving a blue car. Were we observing unwitting memory contamination on a nationwide scale?

Witnesses can be wrong for several reasons. A key reason is that they pick up information from other sources; they combine bits of memory from different experiences. A growing body of research shows that memory more closely resembles a synthesis of experiences than a replay of a videotape[6]. Three decades ago, a method of studying memory distortions was introduced. People watched a simulated crime or accident. Later they were given erroneous information about the details

of the event, such as the false detail that a man had curly rather than straight hair. Many of these people later claimed that they had seen a curly-haired person[7]. Studies such as this showed how leading questions or other forms of misinformation could contaminate the memories of witnesses about events that they had recently experienced[8].

In the past decade, the challenges have become greater. Newer studies showed that you could do more than change a detail here and there in someone's memory. You could actually make people believe that a childhood experience had occurred when in fact it never happened. Examples include being lost in a shopping mall for an extended period of time, being rescued by a lifeguard, or surviving a vicious animal attack[9–12]. How is this possible? In our studies, we enlist family members to help us to persuade their relatives that the events occurred. This method has led about a quarter of our subjects to believe that they were lost in a shopping mall for an extended period of time, and were ultimately rescued by an elderly person and reunited with their families. In other studies, we engaged people in guided imagination exercises. We asked people to imagine for a minute that as a child they had tripped and broken a window with their hand. Later, many of them became confident that the event had occurred. In other studies, we encouraged people to read stories and testimonials about witnessing demonic possession, and even these raised confidence that this rather implausible event had happened.

One recurring issue for memory distortion research is the question of whether the events being reported after such a manipulation might have actually happened. Perhaps the subject did break a window but had forgotten about it — the imagination exercise might have triggered a true memory rather than planting a false one. To prove that false memories can be insinuated into memory by these suggestive techniques, researchers have tried to plant memories that would be highly implausible or impossible. For example, one set of studies asked people to evaluate advertising copy. They were shown a fake print advertisement that described a visit to Disneyland and how they met and shook hands with Bugs Bunny. Later, 16% of these subjects said that they remembered meeting and shaking hands with Bugs Bunny[13]. In follow-up research carried out by Grinley in my laboratory, several presentations of fake advertisments involving Bugs Bunny at Disneyland resulted in 25–35% of subjects claiming to have met Bugs Bunny[14]. Moreover, when these subjects were subsequently asked to report precisely what they remembered about their encounter with Bugs Bunny, 62% remembered shaking his hand and 46% remembered hugging him. A few people remembered touching his ears or tail. One person remembered that he was holding a carrot. The scenes described in the advertisement never occurred, because Bugs Bunny is a Warner Bros. cartoon character and would not be featured at a Disney property.

> "One of the cleverest and most powerful techniques for planting highly implausible false memories involves the use of fake photographs."

Other 'impossible' memories have been recently planted in British students[15]. The false event was "having a nurse remove a skin sample from my little finger." This medical procedure was not one that was carried out in the United Kingdom, according to extensive investigation of health policy records. After guided imagination, many subjects came to remember the non-existent procedure occurring in their childhood. Some embellished their reported memory with significant detail such as, "There was a nurse and the place smelled horrible."

One of the cleverest and most powerful techniques for planting highly implausible false memories involves the use of fake photographs[16]. Subjects were shown a falsified photograph that was made up of a real photograph of the subject and a relative pasted into a prototype photograph of a hot-air balloon (FIG. 1). Family members confirmed that the event had never occurred. Subjects were shown the fake photograph and asked to tell "everything you can remember without leaving anything out, no matter how trivial it may seem." There were two further interviews, and by the end of the series 50% of the subjects had recalled, partially or clearly, the fictitious hot-air balloon ride. Some embellished their reports with sensory details of a hot-air balloon ride during childhood that had never occurred. For example, one subject said "I'm still pretty certain it occurred when I was in sixth grade at, um, the local school there ... I'm pretty certain that mum is down on the ground taking a photo."[16]

These studies, and many more like them, show that people can develop beliefs and memories for events that definitely did not happen to them. They can do this when fed strong suggestions — such as "your family told us about this event" or "look at this photograph of you from childhood". They can even do this when induced to imagine the experiences. Large changes in autobiography can be achieved quickly. Attempts to distinguish the false memories from true ones have occasionally shown statistical differences, such as differences in confidence, vividness or amount of detail[17], or differences in lateralized brain potentials[18,19]. For example, in the hot-air balloon study[16] the real memories were expressed with much more confidence than the fake ones. In most studies, any differences between true and false memories are observed only when comparing large groups of true and false memories, and these differences are typically too small to be useful for classifying a single autobiographical memory report as true or false. Psychological science has not yet developed a reliable way to classify memories as true or false. Moreover, it should be kept in mind that many false memories have been expressed with great confidence.

Implications for society

While researchers continue to investigate false memories, it is evident that there are already lessons to be learned. The fact that the memories of victims and witnesses can be false or inaccurate even though they believe them to be true has important implications for the legal system and for those who counsel or treat victims of crimes.

Some psychotherapists use techniques that are suggestive (along the lines of, "you don't remember sexual abuse, but you have the symptoms, so let's just imagine who might have done it"). These can lead patients to false beliefs and memories, causing great damage to the patients themselves and to those who are accused. In one Illinois case, psychiatrist Bennett Braun was accused by his patient, Patricia Burgus, of using drugs and hypnosis to convince her that she possessed 300 personalities, ate meat loaf made of human flesh and was a high priestess in a satanic cult[20]. By some estimates, thousands of people have been harmed in similar ways by well-meaning providers who apply a 'cure' that ends up being worse than the disease[21]. Law enforcement interrogations that are suggestive can lead witnesses to mistaken memories, even ones that are detailed and expressed with confidence. Hundreds of people have been harmed by witnesses who made a mistake that could have been avoided[22,23]. Of course, even before the police arrive on the scene, witnesses talk to one another and cross contamination can occur. I personally witnessed this when I entered a shop in Cambridge, Massachusetts, moments after a robbery had occurred and before the police arrived. In the immediate aftermath, customers and employees shared

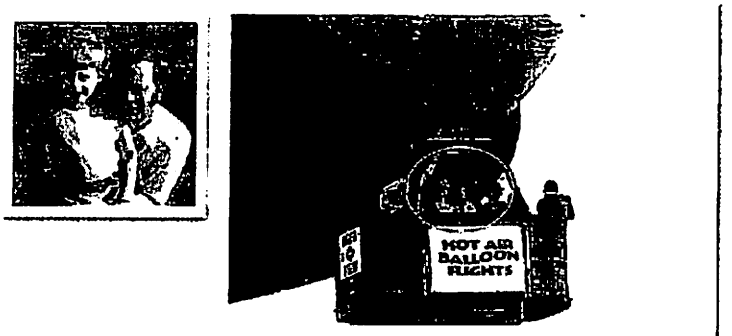


Figure 1 | An example of a composite photograph of a hot-air balloon flight. The photograph on the left was used to create a misleading image (right) that could lead the subject to 'remember' a hot-air balloon flight as a child even when the experience had never occurred. Reproduced, with permission, from REF. 16 © (2002) Psychonomic Society.

their recollections, providing fuel for influencing the thoughts of one another. This is why, during the Washington DC area sniper attacks in 2002, law enforcement officials advised members of the public who might witness the 'next attack' to write down what they saw immediately, even using their hand if they did not have paper. Good advice, but I would suggest having paper handy because the best course of action is to write down everything that can be remembered before witnesses are interrogated or talk to one another. This activity strengthens the memory and protects it to some extent from later contamination[24].

It is often argued that a few false accusations are just the cost of doing business. But this cost includes the potential for the actual perpetrator to commit more crimes, and for the taxpayer to have to pay sizable sums of money in compensation when wrongful convictions are exposed (which probably happens in only a fraction of cases). Although the defendants in most wrongful prosecution cases are government officials or organizations, in one recent case the witness with mistaken memory was successfully sued[25]. Donna Parmeter, a former prison guard, was charged with kidnapping, robbery and torture. She had been identified by the victim, Peter Kretzu, who was tied up, blindfolded and tortured by two masked robbers. Although the attackers wore ski masks, Kretzu claimed that he recognized Donna (from her voice and eyes) and her husband Joseph (from his breathing, laugh, body shape and 'chicken soup' body odour). Kretzu was 100% certain. Donna was eventually exonerated when investigators substantiated her alibi. But she had spent a month in jail, and she later sued, eventually winning a US$100,000 civil judgement against Kretzu. In the past, mistaken witnesses simply went their own ways, although there are a few known instances in which they have made profound apologies to those whom they had falsely accused. Will we now see more cases in which mistaken witnesses end up paying financially for their mistakes?

Although much of the research has focused on wrongful convictions, there is another side to the criminal justice coin. Memory distortions can also contribute to failures to convict a guilty person, not because an innocent person is convicted in their place, but because accurate witness testimony can be undermined. If witnesses misremember some detail, or they are told that their stories conflict with other evidence, they might discount their testimony and be less persuasive than perhaps they should be, or the jury might consider their entire testimony to be unreliable.

Scientific research into memory has the potential to minimize these kinds of problem. Information from psychological scientists (and perhaps neuroscientists) could help to keep the people in power from making decisions on the basis of myths or misconceptions about memory. Scientific knowledge could be shared with relevant individuals in many ways: through workshops for mental health professionals, training for police, seminars for lawyers and judges, judicial instructions or expert testimony for jurors. In one example, Jacob Beard of West Virginia was wrongly convicted of murdering two women and spent many years in prison. He managed to win a second trial. Expert testimony on suggestion and false memory was presented in that second trial, and helped to secure his acquittal. Beard later filed a civil lawsuit, and eventually received a settlement of nearly US$2 million in his case against state and county police[25].

This list of potential venues for education about the nature of memory represents just one proposal for a possible programme for action. Some legislative remedies might also be called for, especially in the most serious cases that can result in a sentence of death. Recently, the Innocence Protection Act was introduced in the United States Congress. It has two useful elements: access to DNA testing for convicted people and improvement in the quality of lawyers who try death penalty cases. Better lawyers might be better acquainted with the problems of memory and how to educate judges and jurors about these problems. Congress will be considering this legislation again in 2003 (REF. 27).

The American Judicature Society proposed the creation of an 'innocence commission' that would study why the legal system has failed in known cases of wrongful conviction. After all, look what the National Transportation Safety Board does when a plane crashes. Few expenses are spared as every aspect of the crash is examined. Not long ago, I proposed an analogous 'National Memory Safety Board' that might concentrate specifically on memory problems that have led to injustice[28]. If the travesties of the past few decades were thoroughly examined side-by-side with scientific knowledge on memory, we would all benefit. It would be too late for the family of Steve Titus, who died of a heart attack at the age of 35 after being falsely convicted of rape. It would be too late for the many death row prisoners who have recently been exonerated by DNA evidence. It would be too late for the scores of innocent defendants who have had to face civil litigation over false claims of satanic ritual abuse and other dubious charges. But it might be in time to keep us from searching for that next white van that does not exist because someone inadvertently planted a false memory.

To reiterate the main points: memory is more prone to error than many people realize. Our memory system can be infused with compelling illusory memories of important events. These grand memory errors have contributed to injustices that could have been avoided or minimized. As a start, I suggest that we all remember an important truth about the mind — paraphrasing Galeano: memory is born anew every day.

Elizabeth Loftus is at 2393 Social Ecology II, University of California, Irvine, California 92697-7085, USA.
e-mail: eloftus@uci.edu

doi:10.1038/nrn1054

1. Duke, L. M., Seltzer, B., Seltzer, J. E. & Vasterling, J. J. Cognitive components of deficit awareness in Alzheimer's disease. *Neuropsychology* 16, 368–368 (2002).
2. Galeano, E. H. translated by Battiega, C. (with Schafer, M.) in *The Book of Embraces* 124–125 (Norton & Co., New York, 1991).
3. Connors, E., Lundregan, T., Miller, N. & McEwan, T. *Convicted by Juries, Exonerated by Science: Case Studies in the Use of DNA Evidence to Establish Innocence After Trial* (National Institute of Justice, Alexandria, Virginia, 1996).
4. Huff, C. R. What can we learn from other nations about the problem of wrongful conviction? *Judicature* 85, 91–97 (2002).
5. Kennedy, H. Beltway sniper notches no. 8. Kills man pumping gas in Va., dodges dragnet. *NY Daily News* [online], (cited 12 October 2002), http://www.nydailynews.com/news/crime_file/story/26905p-24892c.html (2002).
6. Schacter, D. L. *Searching for Memory* (Basic Books, New York, 1996).
7. Loftus, E. F. & Greene, E. Warning: even memory for faces may be contagious. *Law Hum. Behav.* 4, 323–334 (1980).
8. Loftus, E. F. & Hoffman, H. G. Misinformation and memory: the creation of memory. *J. Exp. Psychol. Gen.* 118, 100–104 (1989).
9. Heaps, C. M. & Nash, M. Comparing recollective experience in true and false autobiographical memories. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 920–930 (2001).
10. Loftus, E. F. & Pickrell, J. E. The formation of false memories. *Psychiatr. Ann.* 25, 720–725 (1995).
11. Loftus, E. F. Creating false memories. *Sci. Am.* 277, 70–75 (1997).
12. Porter, S., Birt, A., Yuille, J. C. & Lehman, D. R. Negotiating false memories: interviewer and rememberer characteristics relate to memory distortion. *Psychol. Sci.* 11, 507–510 (2000).
13. Braun, K. A., Ellis, R. E. & Loftus, E. F. Make my memory: how advertising can change our memories of the past. *Psychol. Mark.* 19, 1–23 (2002).
14. Ginley, M. J. *Effects of Advertising on Semantic and Episodic Memory*. Thesis, Univ. Washington (2002).
15. Mazzoni, G. & Memon, A. Imagination can create false autobiographical memories. *Psychol. Sci.* (in the press).
16. Wade, K. A., Garry, M., Read, J. D. & Lindsay, S. A picture is worth a thousand lies. *Psychon. Bull. Rev.* 9, 597–603 (2002).
17. Porter, S., Yuille, J. C. & Lehman, D. R. The nature of real, implanted, and fabricated memories for emotional childhood events. *Law Hum. Behav.* 23, 517–537 (1999).
18. Fabiani, M., Stadler, M. A. & Wessels, P. M. True but not false memories produce a sensory signature in human lateralized brain potentials. *J. Cogn. Neurosci.* 12, 941–949 (2000).
19. Gonsalves, B. & Paller, K. A. Neural events that underlie remembering something that never happened. *Nature Neurosci.* 3, 1316–1320 (2000).
20. Bloomberg, D. Bennett Braun case update: trials set for May, July. *Skeptical Inquirer* 23, 12–13 (1999).
21. Pendergrast, M. *Victims of Memory* 2nd edn (Upper Access, Hinesburg, Vermont, 1996).
22. Huff, C. R., Rattner, A. & Sagarin, E. *Convicted but Innocent: Wrongful Conviction and Public Policy* (Sage Publications, Thousand Oaks, California, 1996).
23. Radelet, M. L. Wrongful convictions of the innocent. *Judicature* 85, 67–68 (2002).
24. Loftus, E. F. *Eyewitness Testimony* (Harvard Univ. Press, Cambridge, Massachusetts, 1996).
25. Pemberton, P. S. Woman falsely accused wins $100,000 judgment. *San Luis Obispo Tribune* B1 (7 December, 2002).
26. Richards, Z. Rainbow case is settled. *Charleston Daily Mail* A1 (4 January 2003).
27. Lage, P. The Innocence Protection Act. *Judicature* 85, 121 (2002).
28. Loftus, E. F. Memory faults and fixes. *Issues Sci. Technol.* 18, 41–50 (2002).

Acknowledgement

I thank the neurophysiologist W. Calvin, for provocative discussions about these issues and general guidance.

Online links

FURTHER INFORMATION

Elizabeth Loftus's homepage:
http://www.sswob.uci.edu/faculty/loftus/
Encyclopedia of Life Sciences: http://www.els.net/
Alzheimer disease | learning and memory
Access to this interactive links box is free online.