AJP

Volume 124 • Number 4 • Winter 2011

The American Journal of Psychology

Edited by Robert W. Proctor *Purdue University*

Book Reviews Dominic W. Massaro *University of California, Santa Cruz*

Obituaries and History of Psychology Alfred H. Fuchs *Bowdoin College*

Pygmalion in the Laboratory

R. ALLEN GARDNER
University of Nevada, Reno

MATTHEW H. SCHEEL
Carroll University

HEIDI L. SHAW
Yakima Valley Community College

Testers and bystanders can inadvertently lead subjects to answers in laboratories and in classrooms, in face-to-face tests of human beings and other animals. Many modern investigators avoid leading by using blind tests scrupulously. This article shows how to design blind tests and illustrates common methodological errors that allow leading to confound experimental results. The object is to help experimenters, editors, and readers detect and avoid a common experimental error that often has profound theoretical implications.

Experiments often need to test human children and nonhuman beings when experimenters and other human beings remain in plain view. This is a common situation in developmental and comparative studies of cognitive function. Pfungst (1907/1965) and Marbe (1917/2011) discovered and described an inherent methodological problem in this testing situation. Testers and bystanders can inadvertently lead subjects to answers. Such leading appears in face-to-face tests of human beings and other animals. Pfungst and Marbe also demonstrated that blind testing eliminates leading by preventing anyone in sight, hearing, touch, or any other contact with the subject from having access to correct answers. Profiting from Pfungst and Marbe, many modern investigators use blind testing scrupulously. Nevertheless, flawed experiments continue to appear in print. This review shows how to design blind tests and illustrates common methodological errors that allow leading to confound experimental results. The object is to help experimenters, editors, and readers detect and avoid a common experimental error that often has profound theoretical implications.

Pygmalion in the Laboratory

In classical mythology, Pygmalion was a sculptor who fell in love with his own creation, the statue of a woman that he named Galatea. Aphrodite, touched

American Journal of Psychology
Winter 2011, Vol. 124, No. 4 pp. 455–461 •

by the beauty of his art and the depth of his love, breathed life into the statue. Pygmalion and Galatea lived happily ever after and even had a son, Paphos. In our time, Rosenthal and Jacobson (1966) coined the expression "Pygmalion in the classroom" for high expectations of teachers that raise classroom performance of some school children and low expectations that lower classroom performance of others. In laboratories as well as in classrooms, biased observers and experimenters can inadvertently carve biased patterns into experimental results in a wide range of experimental contexts.

Pioneering Discoveries

Pygmalion leading can appear in any face-to-face test of human beings or other animals, and this experimental problem has been famous for a century (Marbe, 1917/2011; Pfungst 1907/1965).

Hans was a German horse that seemed to solve problems in simple arithmetic by tapping out numbers with his hoof (Pfungst, 1907/1965). Audiences were amazed, and the name *Clever Hans* became world-famous. Circus trainers and cavalry officers, veterinarians and zookeepers, philosophers and linguists all observed Hans closely, often questioning the horse themselves. Most were highly skeptical, but all failed to discover how a horse could answer correctly without mastering arithmetic. Eventually, an experimental psychologist, Oskar Pfungst (1907/1965), opened the problem to experiment, with a straightforward test. Pfungst whispered one number into Hans's left ear, and von Osten, Hans's owner and trainer, whispered a second number into the horse's right ear. Under that condition, Hans failed. He answered correctly only when he could see someone who already knew the correct answer.

FIGURE 1. Clever Hans, von Osten, and a typical audience. Note the broad-brimmed hats

Alerted by this finding, Pfungst studied face-to-face interactions between von Osten and Hans more closely. He noticed that the trainer usually wore a hat with a large brim. The hat emphasized von Osten's habit of looking down toward Hans's hoof when it was time to start tapping and looking up when it was time to stop tapping. Pfungst soon demonstrated that any questioner could get any number from Hans by lowering and then raising his head, particularly when wearing a hat with a broad brim.

Basso, a young female chimpanzee studied by Karl Marbe in the Frankfurt Zoo, also seemed to solve problems in elementary arithmetic (Marbe, 1917/2011). Asked in German, "How much is 6 plus 4?" Basso selected a tile bearing the number "10," as in Figure 2. Marbe first verified for himself that Basso usually selected the correct tile. Marbe then observed Richard Burkardt, her zookeeper and trainer, as he patiently practiced Basso on simple problems one at a time, as a teacher might instruct a small child in a school room. When asked, Burkardt said he doubted that Basso actually knew arithmetic. Asked how Basso could answer correctly without mastering arithmetic, Burkardt claimed that she used telepathy to read his mind. Accordingly, Marbe tested for telepathy by asking Burkardt to give Basso a series of arithmetic problems and think wrong answers. The result was decisive. Basso repeatedly gave the answer that Burkardt was thinking, regardless of the correct answer in arithmetic. Basso succeeded at telepathy where she failed at arithmetic.

In further tests Marbe showed that when Burkardt questioned Basso he aligned his head and body with the correct tile, whether correct by arithmetic or correct by telepathy. Further testing showed that Basso picked up any tile aligned with Burkardt's head and body. As in the case of von Osten, Burkardt insisted that he never gave any hints whatsoever. With Basso as with Hans, all observers, including Marbe and Pfungst themselves, failed to notice ordinary conversational head movements until they performed systematic experimental tests. Once pointed out, these pragmatic head movements were plain enough for anyone to see clearly. It was only when they were embedded in an engrossing conversational exchange that observers failed to notice them. Arguably, Hans's and Basso's experimentally verified skill at conversational turn-taking has broader implications for comparative intelligence than claims that they might have mastered elementary arithmetic.

Experimental Methods

Where some experimenters have profited from history, others have ignored history at their peril.

BLIND TESTING

Hans depended on informants. He failed when he alone had enough information to answer correctly. The only certain way to prevent Pygmalion leading is to prevent anyone with sight, hearing, touch, or any other conceivable contact from gaining any information that could lead subjects to correct or incorrect answers (Fagot & Leinbach, 1989; Kuhl et al., 2006; Werker & McLeod, 1989).

TESTING INFANTS

In a study of infant speech perception by Kuhl and colleagues (2006), parents held infants in their laps, and both parents and assistants wore headphones that masked speech with music. As part of a larger study investigating child concepts of gender, children in Fagot and Leinbach's (1989) study participated in a gender labeling task. In this task, children selected one of four pictures. During this task, "one experimenter, seated on a low chair facing the child but unable to see the pictures, asked the child to pat, touch, or point to the picture corresponding to the label *boy* or *girl.* A second experimenter controlled the stimulus materials. To avoid cuing the child, the first experimenter was blind to the location of the target picture and order of the pairs, and the second experimenter's face was concealed from the first experimenter by a screen" (p. 665). In one of a set of experiments investigating whether infants showed a preference for male or female infant-directed speech, Werker and McLeod (1989) showed infants a series of audiovisual displays of males and females engaging in infant-directed or adult-directed speech. Throughout testing, parents stood, backs toward the display, holding infants over their shoulders. An experimenter–observer blind to the experimental condition recorded infant gaze direction with regard to the display. Observers and parents alike wore headphones that masked speech with music.

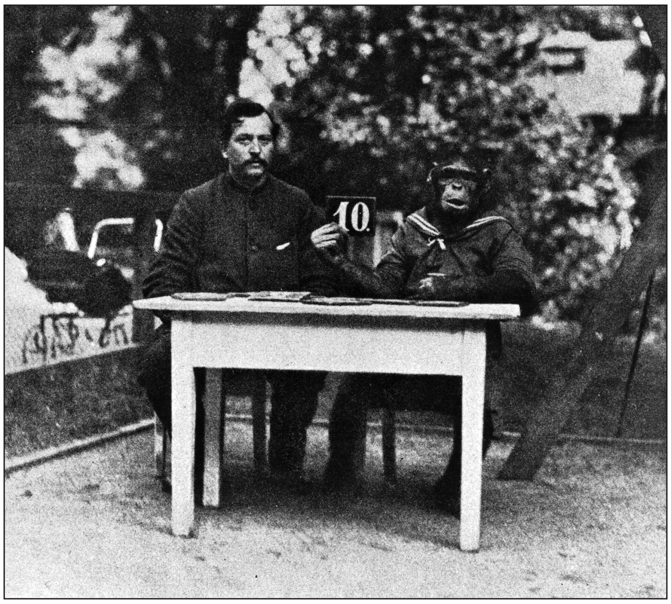

FIGURE 2. Basso sits beside Burkardt and selects a number in answer to his question

VOCABULARY TEST FOR CHIMPANZEES

Gardner and Gardner (1984) designed a blind vocabulary test for chimpanzees that could name pictures in signs of American Sign Language (ASL), the human language most commonly used in deaf communities in North America. The top panel of Figure 3 shows chimpanzee Dar with a human interlocutor and observer (O1). Dar pushes a button (B) to show himself pictures of namable objects on a projection screen (P.S., bottom panel). O1 can see Dar naming pictures but cannot see P.S. and therefore cannot lead Dar to correct answers. The bottom panel shows a second observer (O2) who can see Dar through one-way glass but cannot see P.S. Out of Dar's sight, O2 reports Dar's signs to check the reliability of human readings of Dar's signs and also demonstrates that more than one human observer can read Dar's signs.

An experimenter, also out of Dar's sight, projects slides in unpredictable order, checks procedure (sequence and focus), and receives message slips, directly from O2 and through a message slot (M.S.) from O1. The experimenter serves the essential purpose of allowing O1 and O2 to remain blind to correct answers during testing. Procedures that allow contact

FIGURE 3. (a) O1 can see Dar's signs but cannot see the projection screen. Dar pushes the button (B) to show himself the pictures. (b) O2 can see the chimp through one-way glass but cannot see the projection screen (P.S.). Experimenter checks procedure and receives messages directly from O2 and, via the message slot (M.S.), from O1. Copyright © 1997 by R. Allen Gardner

between experimenters and subjects are less expensive, but they cannot prevent Pygmalion leading.

Under the conditions illustrated in Figure 3, Dar could tell observers (O1 and O2) things they could know only by reading his ASL. Presumably, this is the biological utility of a human language. With a common language, human beings can pass on new information. The better the language, the more the information. The procedure illustrated in Figure 3 tests this basic function. Dar's signs are the only source of information to the observers about the objects on the screen. They cannot lead him because he knows something that they cannot know until he tells them. In this procedure, Dar and O1 must be familiars. Dar tells O1 what he sees on the screen because of the social relationship that he has with his human foster family. They can tell each other new things because they belong to a community with a common language. Gardner and Gardner (1984, 1989) reported high scores with this straightforward procedure for Washoe, Moja, and Tatu, as well as for Dar.

Blocking Communication

Since Pfungst and Marbe, some experimenters have attempted to prevent Pygmalion leading by blocking communication between subjects and testers.

EXHORTATION

Some experimenters have attempted to prevent Pygmalion leading by exhorting testers and others present to refrain from giving hints of any kind (e.g., Cherney, 2003; De Lillo, Spinozzi, Truppa, & Naylor, 2005; Kikuno, Mitchell, & Ziegler, 2007). For example, De Lillo et al. wrote that "both experimenters were extremely careful in avoiding giving spurious cues by means of their body posture or the direction of their gaze during the trials" (p. 158). This popular procedure is easy and cheap, but it is also unlikely to succeed. Both von Osten and Burkardt insisted that they never gave any hints to Hans or to Basso. Pfungst and Marbe described both trainers as sincere and honest. Moreover, many skeptical experts traveled long distances for the chance to test Hans, with the express purpose of proving that a horse could never master elementary arithmetic. Only systematic experiments demonstrated that honest, even hostile questioners were inadvertently leading Hans and Basso, in spite of themselves.

Often mothers or close relatives and friends accompany young children, and very young children sit in their mother's lap during testing (Cashon & Cohen, 2000; Herrmann, Call, Hernández-Lloreda, Hare, & Tomasello, 2007; Surian, Caldi, & Sperber, 2007). For example, Cashon and Cohen wrote that infants "sat on a parent's lap facing the video monitor. . . . Parents were instructed not to talk, point or interact in any way that would influence or distract the infant" (p. 436). Surian et al. instructed mothers "to be silent and avoid any interference with their babies during the experiment" (pp. 582–583). Similarly, in Herrmann et al., children were "accompanied by a parent who was told not to influence or help in any way" (p. 1362). Exhorting mothers to refrain from empathy with their infants seems insensitive to mother–infant relations in most cultures. If a parent must be present to ensure a child's cooperation, practical, effective procedures can prevent parental access to test stimuli (Fagot & Leinbach, 1989; Kuhl et al., 2006; Werker & McLeod, 1989).

OBSCURING EYES

In another popular but ineffective procedure, experimenters have obscured the eyes of testers, as if Pygmalion leading depended on eye movements alone. In some cases, testers have worn baseball caps with long bills to cover their eyes (e.g., Call, 2001; Call & Rochat, 1997; Suda-King, 2008). Still others have bowed their heads and looked down (e.g., Okamoto-Barth & Call, 2008). However, Pfungst reported, repeatedly and in great detail, that Hans succeeded when von Osten, Pfungst, or any other interlocutor wore a hat with a broad brim that obscured their eyes as they lowered their heads to look down toward Hans's hoof. Von Osten's broad brimmed hat is emblematic in Pfungst's descriptions of von Osten testing Hans (see Figure 1). Hats with broad brims exaggerate head orientation even as they obscure eye movement. Marbe demonstrated repeatedly that Basso followed head orientation to arithmetically correct as well as telepathically correct numbers even when her human interlocutor shut his eyes.

Leading Up and Leading Down

Both Pfungst and Marbe showed that Pygmalion leading can introduce spuriously low scores as well as spuriously high scores. Low scores can confirm negative expectations just as high scores can confirm positive expectations. When Basso chose the numbers that her teacher was thinking instead of the correct answers to his voiced questions, she confirmed his belief in her telepathic powers. Only Marbe's systematic experiments revealed that Basso was following Burkardt's inadvertent head and body orientation, both when her arithmetic was correct and when her telepathy was correct. Von Osten taught Hans short division because he believed that long division was too difficult for a horse. To von Osten, Hans's failure at tests of long division validated Hans's success at short division. Pfungst's experimental demonstrations never convinced von Osten that the trainer, or anyone else, was leading Hans to answers.

Discussions of "Clever Hans errors" (Gray, 2007, pp. 26–28, 43–44) often imply that Pygmalion leading leads subjects only to correct answers. Therefore, some experimenters have ruled out any possibility of leading when subjects do not score above chance, as in De Lillo et al. (2005). "The fact that children were not able to perform above chance level in these trials allows us to rule out the possibility that they were solving the task on the basis of spurious cues involuntarily conveyed by the testers" (p. 160). This common misunderstanding can be avoided by using the term *Pygmalion leading* for a pattern of correct and incorrect answers, now favoring one species, now another, now one age group or ethnic group, now another, depending on expectations of testers. Without appropriate procedures, perfectly innocent experimenters can shape results either up or down, as Pygmalion shaped Galatea from raw ivory.

Valid Comparison

Increasingly, claims of evidence for cognitive gaps between human beings and other animals fail to prevent Pygmalion leading and also fail to provide comparable testing conditions. To support their title "Darwin's Mistake," for example, Penn, Holyoak, and Povinelli (2008) cite only experiments that confound species comparisons with opportunities for Pygmalion leading as well as biased testing conditions. Video published in *Science* to supplement Herrmann et al. (2007) and Warneken, Chen, and Tomasello (2006) shows how marked differences between testing conditions prohibit any valid experimental conclusions about differences between species. Differences in experimenter expectations and rapport between experimenters and subjects are abundantly illustrated in this published video. Chimpanzees take their tests alone in a bare, dark cage. An observer–tester presents testing materials from outside the cage with minimum social interaction. Children, on the other hand, take their tests in a cheery, decorated nursery school room with one or two familiar adult companions (possibly relatives or friends), and a very sociable experimenter presents testing materials. Similar differences in testing conditions appear clearly in still photos and line drawings published with laboratory reports. Quite apart from positive and negative leading, differences in testing conditions could easily account for differences between children and chimpanzees in whole programs of research.

In Povinelli and Eddy (1996), captive chimpanzees failed to discriminate between an ordinary human being and one with a bucket over his head, or even to learn this discrimination after repeated trials. Povinelli and Eddy reported that "not only were the subjects' scores low in the first session, but they declined further in the second session" (p. 41). "Given that their performance was at chance in most treatment conditions . . . it is difficult to make the case that the subjects were relying on such [Pygmalion] cues" (pp. 47–48). Povinelli and Eddy attributed this failure to an evolutionary gap between human and chimpanzee cognition. Meanwhile, observers of chimpanzees in the wild regularly report reconciliation (Wittig & Boesch, 2005), cooperative hunting (Boesch, 2001), memory for food locations (Normand, Ban, & Boesch, 2009), distinctions between familiar and unfamiliar vocalizations (Herbinger, Papworth, Boesch, & Zuberbühler, 2009), and innovation (Kummer & Goodall, 2003). Results reported in Povinelli and Eddy probably tell more about the cognitively and socially deprived laboratory conditions in which they reared and maintained their chimpanzee subjects than they can tell us about any evolutionary gap. Also, without effective procedures to prevent Pygmalion leading, biased testers may have inadvertently led chimpanzees to incorrect choices.

Summary

Pragmatic devices in ordinary human conversation can lead subjects to answers, without any awareness on the part of questioners. Often questioners insist that they cannot possibly be leading subjects in any way. Nevertheless, appropriate tests show that scrupulously honest questioners, and often deeply skeptical questioners, can lead subjects to incorrect as well as to correct answers. Pygmalion leading appears in laboratories as well as in classrooms in face-to-face tests of human beings and other animals. Pervasive as Pygmalion leading may be in common testing situations, straightforward procedures can eliminate it entirely.

NOTE

Address correspondence about this article to R. Allen Gardner, University of Nevada, Reno, NV 89557 (e-mail: gardner@unr.edu).

REFERENCES

Boesch, C. (2001). Cooperative hunting roles among Taï chimpanzees. *Human Nature, 13*(1), 27–46.

Call, J. (2001). Object permanence in orangutans (*Pongo pygmaeus*), chimpanzees (*Pan troglodytes*), and children (*Homo sapiens*). *Journal of Comparative Psychology, 115*(2), 159–171.

Call, J., & Rochat, P. (1997). Perceptual strategies in the estimation of physical quantities by orangutans (*Pongo pygmaeus*). *Journal of Comparative Psychology, 111*(4), 315–329.

Cashon, C., & Cohen, L. (2000). Eight-month-old infants' perception of possible and impossible events. *Infancy, 1*(4), 429–446.

Cherney, I. (2003). Young children's spontaneous utterances of mental terms and the accuracy of their memory behaviors: A different methodological approach. *Infant and Child Development, 12*(1), 89–105.

De Lillo, C., Spinozzi, G., Truppa, V., & Naylor, D. M. (2005). A comparative analysis of global and local processing of hierarchical visual stimuli in young children (*Homo sapiens*) and monkeys (*Cebus apella*). *Journal of Comparative Psychology, 119*(2), 155–165.

Fagot, B. I., & Leinbach, M. D. (1989). The young child's gender schema: Environmental input, internal organization. *Child Development, 60*(3), 663–672.

Gardner, B. T., & Gardner, R. A. (1989). Prelinguistic development of children and chimpanzees. *Human Evolution, 4*(6), 433–460.

Gardner, R. A., & Gardner, B. T. (1984). A vocabulary test for chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 98*(4), 381–404.

Gray, P. (2007). *Psychology* (5th ed.). New York, NY: Worth Publishers.

Herbinger, I., Papworth, S., Boesch, C., & Zuberbühler, K. (2009). Vocal, gestural and locomotor responses of wild chimpanzees to familiar and unfamiliar intruders: A playback study. *Animal Behaviour, 78*(6), 1389–1396.

Herrmann, E., Call, J., Hernández-Lloreda, M., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317*(5843), 1360–1366.

Kikuno, H., Mitchell, P., & Ziegler, F. (2007). How do young children process beliefs about beliefs?: Evidence from response latency. *Mind & Language, 22*(3), 297–316.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13-F21.

Kummer, H., & Goodall, J. (2003). Conditions of innovative behaviour in primates. In S. M. Reader & K. N. Laland (Eds.), *Animal innovation* (pp. 223–235). New York, NY: Oxford University Press.

Marbe, K. (1917). Die Rechenkunst der Schimpansen Basso im Frankfurter zoologischen Garten nebst Bemerkungen zur Tierpsychologie und einem offenem Brief an Herrn Krall. *Fortschritte der Psychologie und Ihren Anwendungen, 4,* 135–187.

Marbe, K. (2011). Arithmetic in the chimpanzee Basso in the Frankfurt Zoological Park together with remarks to animal psychology and an open letter to Herr Krall (H. L. Shaw, Trans.). *The American Journal of Psychology, 124,* 463–488. (Original work published 1917).

Normand, E., Ban, S. D., & Boesch, C. (2009). Forest chimpanzees (*Pan troglodytes verus*) remember the location of numerous fruit trees. *Animal Cognition, 12*(6), 797–807.

Okamoto-Barth, S., & Call, J. (2008). Tracking and inferring spatial rotation by children and great apes. *Developmental Psychology, 44*(5), 1396–1408.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31*(2), 109–130.

Pfungst, O. (1965). *Clever Hans (The horse of Mr. Von Osten). A contribution to experimental animal and human psychology* (C. L. Rahn, Trans.). New York, NY: Holt. (Original work published 1907)

Povinelli, D. J., & Eddy, T. (1996). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development, 61*(3), v–vi, 1–152.

Rosenthal, R., & Jacobson, L. (1966). Teachers' expectancies: Determinants of pupils' IQ gains. *Psychological Reports, 19*(1), 115–118.

Suda-King, C. (2008). Do orangutans (*Pongo pygmaeus*) know when they do not remember? *Animal Cognition, 11*(1), 21–42.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*(7), 580–586.

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development, 77*(3), 640–663.

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A development study of attentional and affective responsiveness. *Canadian Journal of Psychology, 43*(2), 230–246.

Wittig, R., & Boesch, C. (2005). How to repair relationships: Reconciliation in wild chimpanzees (*Pan troglodytes*). *Ethology, 111*(8), 736–763.

Library Recommendation Form

THE
American Journal
OF
Psychology

To: Librarian / Library Acquisition Committee

From: ______________________________________

Position: ____________________ **Dept:** ___________

Email: ______________________

Phone: ___________________

I recommend that the library subscribe to *The American Journal of Psychology*. Please include this journal in your next serials review meeting with my recommendation to subscribe.

Signature: __________________________________

Date: _________

For subscription information please contact the University of Illinois Press online at http://www.press.uillinois.edu/journals/ajp.html