

*This copy is for your personal, non-commercial use only.*

**If you wish to distribute this article to others**, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

**Permission to republish or repurpose articles or portions of articles** can be obtained by following the guidelines [here](#).

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of August 17, 2014 ):**

A correction has been published for this article at:

<http://www.sciencemag.org/content/317/5839/750.full.html>

<http://www.sciencemag.org/content/314/5799/595.1.full.html>

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/311/5763/1020.full.html>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/content/suppl/2006/02/13/311.5763.1020.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/311/5763/1020.full.html#related>

This article has been **cited by** 42 article(s) on the ISI Web of Science

This article has been **cited by** 13 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/311/5763/1020.full.html#related-urls>

This article appears in the following **subject collections**:

Neuroscience

<http://www.sciencemag.org/cgi/collection/neuroscience>

5. S. A. Thomas, A. M. Matsumoto, R. D. Palmiter, *Nature* **374**, 643 (1995).
6. T. M. Tzschentke, *Prog. Neurobiol.* **56**, 613 (1998).
7. J. R. Schank *et al.*, *Neuropsychopharmacology*, published online 14 December 2005 (10.1038/sj.npp.1301000).
8. R. Ventura, A. Alcaro, S. Puglisi-Allegra, *Cereb. Cortex* **15**, 1877 (2005).
9. C. Drouin, G. Blanc, F. Trovero, J. Glowinski, J. P. Tassin, *Neuroreport* **12**, 3483 (2001).
10. C. Drouin *et al.*, *J. Neurosci.* **22**, 2873 (2002).
11. T. S. Hnasko, B. N. Sotak, R. D. Palmiter, *Nature* **438**, 854 (2005).
12. J. M. Delfs, Y. Zhu, J. P. Druhan, G. Aston-Jones, *Nature* **403**, 430 (2000).
13. V. G. Olson, R. D. Palmiter, data not shown.
14. R. Y. Moore, F. E. Bloom, *Annu. Rev. Neurosci.* **2**, 113 (1979).
15. J. M. Delfs, Y. Zhu, J. P. Druhan, G. S. Aston-Jones, *Brain Res.* **806**, 127 (1998).
16. P. M. Milner, *Can. J. Psychol.* **45**, 1 (1991).
17. M. R. Zarrindast, T. Bahreini, M. Adl, *Pharmacol. Biochem. Behav.* **73**, 941 (2002).
18. H. Sahraei *et al.*, *Pharmacol. Biochem. Behav.* **78**, 135 (2004).
19. K. Nader, D. van der Kooy, *Behav. Neurosci.* **110**, 389 (1996).
20. M. J. Christie, J. T. Williams, P. B. Osborne, C. E. Bellchambers, *Trends Pharmacol. Sci.* **18**, 134 (1997).
21. M. Spreng, S. Cotecchia, F. Schenk, *Neurobiol. Learn. Mem.* **75**, 214 (2001).
22. J. Knauber, W. E. Muller, *Eur. Neuropsychopharmacol.* **10**, 423 (2000).
23. A. Auclair, C. Drouin, S. Cotecchia, J. Glowinski, J. P. Tassin, *Eur. J. Neurosci.* **20**, 3073 (2004).
24. A. F. Arnsten, B. M. Li, *Biol. Psychiatry* **57**, 1377 (2005).
25. M. L. Laorden, M. T. Castells, M. V. Milanés, *Br. J. Pharmacol.* **136**, 67 (2002).
26. T. Watanabe *et al.*, *Psychopharmacology* **170**, 80 (2003).
27. X. Wang, X. Cen, L. Lu, *Eur. J. Pharmacol.* **432**, 153 (2001).
28. Y. Shaham, D. Highfield, J. Delfs, S. Leung, J. Stewart, *Eur. J. Neurosci.* **12**, 292 (2000).
29. J. Grenhoff, M. Nisell, S. Ferre, G. Aston-Jones, T. H. Svensson, *J. Neural Transm. Gen. Sect.* **93**, 11 (1993).
30. We thank W. Curtis for maintaining the DBH mouse colony, Sumitomo Pharmaceuticals for DOPS, C. Hutter for statistical advice, P. M. Tran for technical assistance, and our colleagues for constructive suggestions. Supported by an NIH National Research Service Award grant (V.G.O.) and an institutional Program Project grant from the National Institute on Drug Abuse.

### Supporting Online Material

www.sciencemag.org/cgi/content/full/311/5763/1017/DC1  
Materials and Methods  
Figs. S1 to S3

24 August 2005; accepted 4 January 2006  
10.1126/science.1119311

## Causal Reasoning in Rats

Aaron P. Blaisdell,<sup>1\*</sup> Kosuke Sawa,<sup>2</sup> Kenneth J. Leising,<sup>1</sup> Michael R. Waldmann<sup>3</sup>

Empirical research with nonhuman primates appears to support the view that causal reasoning is a key cognitive faculty that divides humans from animals. The claim is that animals approximate causal learning using associative processes. The present results cast doubt on that conclusion. Rats made causal inferences in a basic task that taps into core features of causal reasoning without requiring complex physical knowledge. They derived predictions of the outcomes of interventions after passive observational learning of different kinds of causal models. These competencies cannot be explained by current associative theories but are consistent with causal Bayes net theories.

The ability to acquire and reason with causal knowledge is among our most central human cognitive competences (1). Causal knowledge serves two important functions: It allows us to predict outcomes on the basis of observations, and it underlies our ability to control events in the world. We investigated whether animals understand the relation between observations and interventions, which some philosophers regard as a core feature of causal reasoning (2–4).

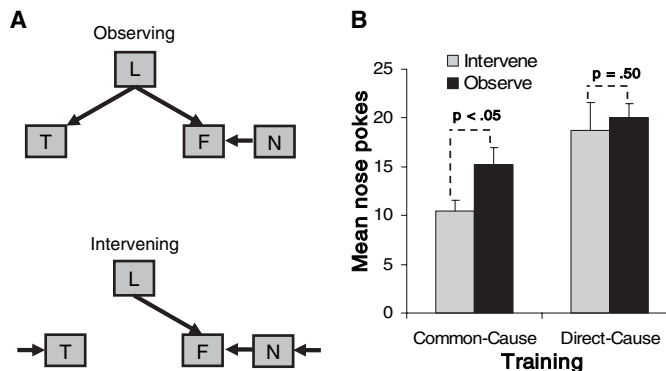
Although a number of psychologists have claimed that both humans and animals use basic associative mechanisms to learn about causal relations (5), human studies have demonstrated a deeper understanding of causal relations that cannot be reduced to associative learning (6–8). In contrast, research on the cognitive competencies of nonhuman primates concludes that they demonstrate a superficial understanding of the association between tool use and its effects but fail to comprehend the unobservable physical mechanisms underlying these relations [(9–11), but see (12, 13)]. It may well be, however, that nonhuman animals lack knowledge about physical mechanisms but still are capable

of basic causal reasoning. The capacity to derive predictions for interventions after purely observational learning is a core competency that is not reducible to associative learning (14).

Humans and animals can learn associations between passively observed events (Pavlovian conditioning) as well as between interventions and outcomes (instrumental conditioning). Moreover, these two learning modes may interact (15). An understanding of the interrelations between observations (“seeing”) and interventions (“doing”), however, requires more

sophisticated representations. Simple transfer from observational learning can lead to inadequate predictions for interventions. For example, barometer readings statistically predict the weather, but at the same time, setting the barometer to an arbitrary reading does not influence the weather. Both relations could be learned with associative mechanisms in separate observational and instrumental learning trials, but associative theories are incapable of deriving correct predictions for interventions after observational learning when no prior instrumental learning is available.

The causal model in Fig. 1A shows how predictions for interventions can be derived from observations. Imagine that an animal learns in an observational Pavlovian learning phase that a light cue (L) temporally precedes both a tone stimulus (T) and food (F), thus learning a common-cause model with two effects (top panel). After learning this model, observing T should, via L, lead to the predictive inference that F should also be present. However, if the animal learns in the test phase that a newly introduced lever turns on T, it should be more



**Fig. 1. (A)** Causal model used in experiment 1. L (light) is the common cause of T (tone) and F (food); N (noise) is the direct cause of F. (Top panel) Observed causal relations. (Bottom panel) Model modified under the assumption of an intervention in T and N. **(B)** Experiment 1: Mean nose pokes in response to test stimulus T ( $P < 0.05$ ) in the

common-cause condition and to N ( $P > 0.50$ ) in the direct-cause condition after a lever press (intervene) or no lever press (observe). Bars indicate SEM. Planned comparisons from a two-way mixed analysis of variance (ANOVA) are shown. There was a main effect of causal model (common or direct),  $F(1, 21) = 6.01$ ,  $P < 0.05$ , and an interaction between causal model and test condition (intervene or observe),  $F(1, 21) = 4.31$ ,  $P = 0.05$ .

<sup>1</sup>Department of Psychology, University of California, Los Angeles, CA 90095, USA. <sup>2</sup>Japan Society for the Promotion of Science, Nagoya University, Nagoya 464-8601, Japan. <sup>3</sup>Department of Psychology, University of Göttingen, 37073 Göttingen, Germany.

\*To whom correspondence should be addressed. E-mail: blaisdell@psych.ucla.edu

reluctant to predict F (bottom panel). Generating T by means of an alternative cause—the lever—does not predict F because the manipulation of an effect does not influence its cause (L). A dissociation between seeing and doing would be remarkable, because in the observational learning phase T is positively correlated with L.

The only theoretical model that derives correct predictions for interventions from observational learning data is causal Bayes nets (2–4). Predictions for observations make use of the full causal model acquired during observational learning (top panel). Predictions for interventions, however, are based on a modified graph (bottom panel); the insight that generating T in the common-cause model happens independently of its usual cause L is modeled by removing the causal arrow that leads into the manipulated effect: a manipulation called graph surgery (3). Because the manipulated T is unrelated to L, the likelihood of L's other effect F should not be altered by T's presence.

A possible alternative associationist explanation of the failure to expect F after an intervention in T may be that the animal does not expect F because it lacks prior instrumental learning experiences relating lever presses to F. This alternative theory, however, erroneously also predicts a failure to expect F in the presence of noise (N), after these events had been paired during observational learning (Fig. 1A). Because of the direct causal link between N and F, causal Bayes nets predict that animals should equally expect F, regardless of whether N is observed or generated by an intervention. Recent research with similar tasks has shown (14) that human participants are capable of deriving correct predictions for interventions on the basis of observational data (16).

In experiment 1, 32 rats were trained on the causal model shown in Fig. 1A, using an observational Pavlovian procedure (17). Training consisted of three types of trials interspersed within each session. The first type of trial was presentations of stimulus L (a 10-s flashing light or click train) forward-paired with

stimulus T (a 10-s tone or noise); the second was presentations of stimulus L forward-paired with stimulus F (a 10-s delivery of sucrose solution); the third was simultaneous presentations of stimulus N (a 10-s noise or tone) and 10 s of F. We trained each causal link in the common-cause model separately to make it more likely that subjects did not induce a direct link between effects T and F.

Why did the rats not induce that the alternative effect is always absent when the cause and one effect are present (that is, conditioned inhibition)? With few learning trials, rats tend to integrate individual learning relations into a coherent integrated model. Only after many trials do rats encode the explicit absence of the nonpresented cues (18). Supporting these findings, the results of all our experiments show that rats induced second-order excitatory rather than inhibitory relations (19). Apparently, in the initial phases of learning, rats tend to conservatively treat the absent but expected events as possibly present but missed. A similar ability to combine individually learned causal links into complex causal models has been demonstrated in humans (20).

Do rats treat L as a common cause of both T and F, and do they correctly differentiate between seeing and doing with respect to T and N? Rats were allocated to one of four test conditions and were placed in the conditioning chamber with a lever present. This lever had not been present in the observational learning phase, so that no prior instrumental knowledge was available. Rats in condition intervene-T received a 10-s presentation of T each time they pressed the lever. Rats in condition observe-T merely observed presentations of T independently of any emitted lever presses. Conditions intervene-N and observe-N were conducted in an identical fashion, except that N was either the product of an intervention by lever pressing or was observed. We recorded the number of nose pokes into the magazine where F had been delivered during the training phase, to assess the rats' expectation of F.

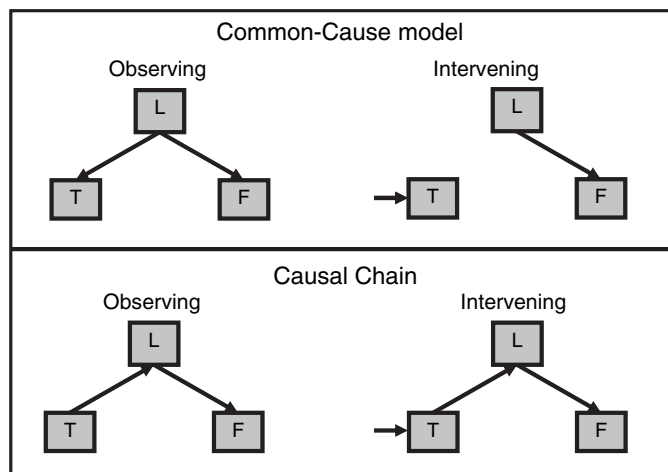
Causal Bayes nets predict that observing T in condition observe-T should lead the rats to reason that the temporally prior cause L was probably present (but missed), and to consequently expect that F should also be present; therefore, they should emit many nose pokes. In contrast, rats in condition intervene-T should attribute T to their intervention and therefore expect L and consequently its effect, F, to occur with the probability corresponding to the base rate of its cause L. Consequently, we should observe a lower rate of nose poking in condition intervene-T than in condition observe-T. There should not be any difference in rates of nose poking, however, between conditions intervene-N and observe-N. The direct causal relationship should lead the rats to expect F regardless of whether N was observed or intervened on at test. Unlike causal Bayes nets, associationist theories predict equivalent nose poking in the presence of T in both the observe and intervene conditions.

Figure 1B shows the mean rate of nose poking per 10-s presentation of stimuli T and N as a function of test condition (with a maximum rate of 100 nose pokes per presentation). As predicted by causal Bayes nets, rats that produced T through a lever-press intervention (condition intervene-T) made fewer nose pokes than rats that merely observed T (condition observe-T). However, rats that intervened in N (condition Intervene-N), which was trained as a direct predictor of F, did not nose poke less than rats that merely observed N (condition observe-N). [An analysis of the lever press data ruled out selective interference between lever pressing and nose poking (17).]

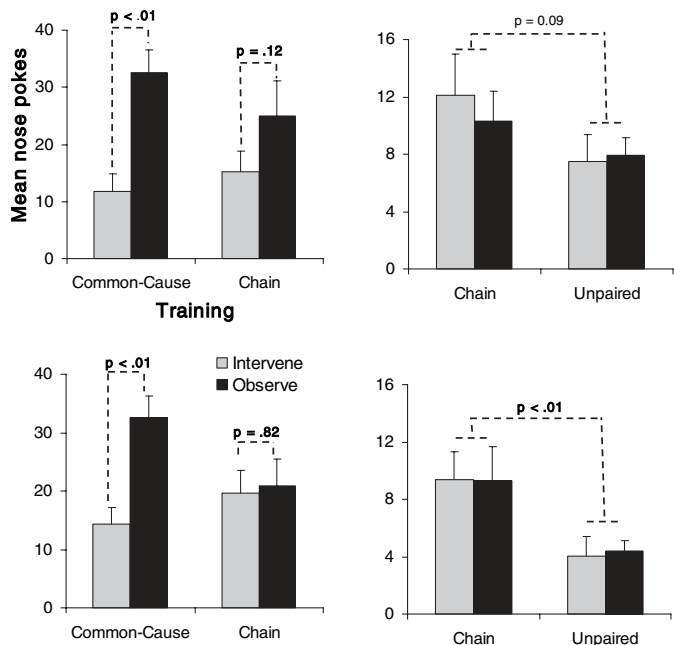
In experiment 1, we observed a dissociation between seeing and doing within the common-cause model, whereas both tasks led to identical expectations with the direct causal link, which is consistent with causal Bayes nets. A critic might point out that we found a dissociation within a complex causal model with two separately learned links (the common-cause model), whereas we found similar responses to the less complex direct link. To rule out complexity or second-order learning as the basis of our dissociation, we compared a common-cause condition with an equally complex causal chain in which the individual causal links were also presented separately (that is, second-order conditioning) (Fig. 2). Whereas causal Bayes nets predict a dissociation between seeing and doing in the common-cause model, no such dissociation is expected for the causal chain. Regardless of whether the initial cause (T) of the chain is observed or generated by means of an intervention, the intermediate (L) and final effect (F) should equally be expected.

In experiment 2a, rats received either common-cause training, as in experiment 1, or causal-chain training, which was identical except that T preceded L during observational learning (17). In the test phase, groups

**Fig. 2.** Common-cause and causal chain models from experiment 2. **(Left)** Observed causal relations. **(Right)** Model modified under the assumption of an intervention in T.



**Fig. 3.** Experiment 2a (left panels): Mean nose pokes during test stimulus T (top panel) or 10 s after the termination of T (bottom panel) after a lever press (intervene;  $P = 0.01$  in both panels) or no lever press (observe;  $P = 0.12$  and  $0.82$  in top and bottom panels, respectively). Common-cause and chain indicate the type of causal model training. Bars indicate SEM. Planned comparisons from two-way ANOVAs are shown. Experiment 2b (right panels): Mean nose pokes during test stimulus T (top panel) or 10 s after the termination of T (bottom panel) after a lever press (intervene) or no lever press (observe). Chain and unpaired indicate the type of causal model training. Bars indicate SEM.  $P = 0.09$  and  $0.01$  in top and bottom panels, respectively, for the main effect of training.



common-cause-intervene and chain-intervene received presentations of T each time the lever was pressed. Groups common-cause-observe and chain-observe merely observed T. We report the number of nose pokes during the 10-s presentation of T and during the 10-s period beginning 10 s after the termination of T (post-T interval 2) for all subjects. In the chain condition, F should rationally be expected between 10 and 20 s after delivery of T (19). In contrast, the expected time of delivery of F for rats that received common-cause training is during T itself.

Figure 3 shows the mean rate of nose poking on test trials with T. Group common-cause-intervene nose poked less than group common-cause-observe, which replicates the pattern of experiment 1. In contrast, no difference was found between groups chain-intervene and chain-observe, as predicted by causal Bayes nets.

Rats in group chain-intervene did not nose poke more than did rats in group common-cause-intervene. This low level of responding

does not reflect a failure to learn a causal chain, however. Experiment 2b replicated the chain condition and added groups for which T and L were unpaired during observational learning (17). Figure 3 reveals no difference between seeing and doing, as predicted by causal Bayes nets. Moreover, responding in the causal-chain groups was higher than in the unpaired groups, which signifies that the rats had indeed learned the second-order chain relations.

A number of researchers have recently concluded that causal reasoning is a faculty that divides humans from animals (7, 9–11). The present results cast doubt on that conclusion. With tasks that did not require complex physical knowledge, the experiments have shown that rats grasp the relationship between seeing and doing. Rats made correct inferences for instrumental actions on the basis of purely observational learning, and they correctly differentiated between common-cause models, causal chains, and direct causal links. These results contradict the view that causal learning in rats is solely driven by associative learn-

ing mechanisms, but they are consistent with causal Bayes net theories. The core competency of reasoning with causal models seems to be already in place in animals, even when elaborate physical knowledge may not yet be available.

### References and Notes

- D. Hume, *Treatise on Human Nature*, L. A. Selby-Bigge, Ed. (Oxford Univ. Press, London, 1964) (first published 1739).
- P. Spirtes, C. Glymour, R. Scheines, *Causation, Prediction, and Search* (Springer-Verlag, New York, 1993).
- J. Pearl, *Causality* (Cambridge Univ. Press, Cambridge, 2000).
- J. Woodward, *Making Things Happen* (Oxford Univ. Press, Oxford, 2003).
- A. Dickinson, D. Shanks, *Causal Cognition*, D. Sperber, A. J. Premack, Eds. (Clarendon Press, Oxford, 1995), pp. 5–25.
- A. Gopnik et al., *Psychol. Rev.* **111**, 3 (2004).
- A. Gopnik, L. Schulz, *Trends Cogn. Sci.* **8**, 371 (2004).
- M. R. Waldmann, *J. Exp. Psychol. Learn. Mem. Cogn.* **26**, 53 (2000).
- M. Tomasello, J. Call, *Primate Cognition* (Oxford Univ. Press, London, 1997).
- D. J. Povinelli, *Folk Physics for Apes* (Oxford Univ. Press, New York, 2000).
- E. Visalberghi, L. Limongelli, *J. Comp. Psychol.* **108**, 15 (1994).
- N. J. Mulcahy, J. Call, R. I. M. Dunbar, *J. Comp. Psychol.* **119**, 23 (2005).
- L. R. Santos, C. T. Miller, M. D. Hauser, *Anim. Cogn.* **6**, 269 (2003).
- M. R. Waldmann, Y. Hagmayer, *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 216 (2005).
- A. Dickinson, B. Balleine, in *Steven's Handbook of Experimental Psychology*, vol. 3, H. Pashler, R. Gallistel, Eds. (Wiley, New York, ed. 3, 2002), pp. 497–533.
- S. A. Sloman, D. A. Lagnado, *Cogn. Sci.* **29**, 5 (2005).
- Materials, methods, and procedural details are available as supporting material on *Science Online*.
- H. Yin, R. C. Barnett, R. R. Miller, *J. Exp. Psychol. Anim. Behav. Process.* **20**, 419 (1994).
- H. I. Savastano, R. R. Miller, *Behav. Process.* **44**, 147 (1998).
- J. C. Perales, A. Catena, A. Maldonado, *Learn. Motiv.* **35**, 115 (2004).
- We thank T. Dickinson, K. Holyoak, K. Leising, W. D. Stahlman, T. Beckers, and D. Penn for helpful comments. This research was supported by National Institute of Mental Health grant MH12531 (A.P.B.) and by the Japan Ministry of Science (K.S.).

### Supporting Online Material

www.sciencemag.org/cgi/content/full/311/5763/1020/DC1  
Materials and Methods  
References

26 October 2005; accepted 11 January 2006  
10.1126/science.1121872

## ERRATUM

Post date 27 October 2006

**Reports:** "Causal reasoning in rats" by A. P. Blaisdell *et al.* (17 Feb. 2006, p. 1020). The wrong input data were used to generate Fig. 1B. The corrected figure is shown here. The error does not change the conclusions of the paper.

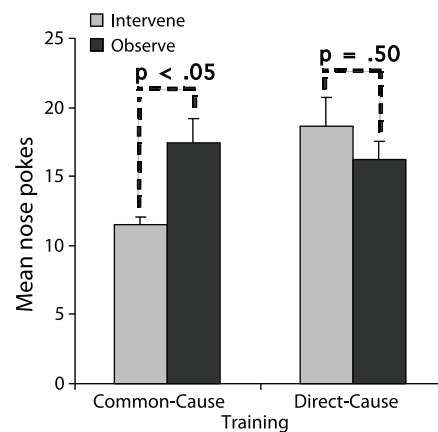


Fig. 1B

# ERRATUM

Post date 10 August 2007

**Reports:** "Causal reasoning in rats" by A. P. Blaisdell *et al.* (17 February 2006, p. 1020). There are three minor typos in the Supporting Online Materials. First, test sessions for Experiments 1, 2a, and 2b were 60 minutes. Second, the number of background nose pokes in Experiment 1 were  $2793 \pm 571$  (Conditions Intervene-T),  $3051 \pm 991$  (Condition Observe-T),  $2885 \pm 823$  (Condition Intervene-N), and  $2849 \pm 514$  (Condition Observe-N). Third, in Experiment 1, the  $F$  value for the planned comparison between condition Intervene-T versus Observe-T was 9.07,  $p < 0.05$ . A reanalysis of the data from Experiment 2b revealed that the test-trial data for one subject from group Unpaired-Observe was inadvertently counted twice in the statistical analyses. A reanalysis on the corrected data results in a change of three  $F$  values by a tenth of a point or less, and thus has no effect on the outcome of the analyses. The authors failed to revise the caption in the corrected Fig. 1 [*Science* **314**, 595 (2006)]. The  $F$  values were slightly different than those reported. Corrected values show the main effect of inference type =  $F(1, 21) = 4.57$ ,  $P < 0.05$ , and the interaction =  $F(1, 21) = 5.69$ ,  $P < 0.05$ .