Effects of Correlational Strength and Correlational Indeterminacy on Judgments of Causality

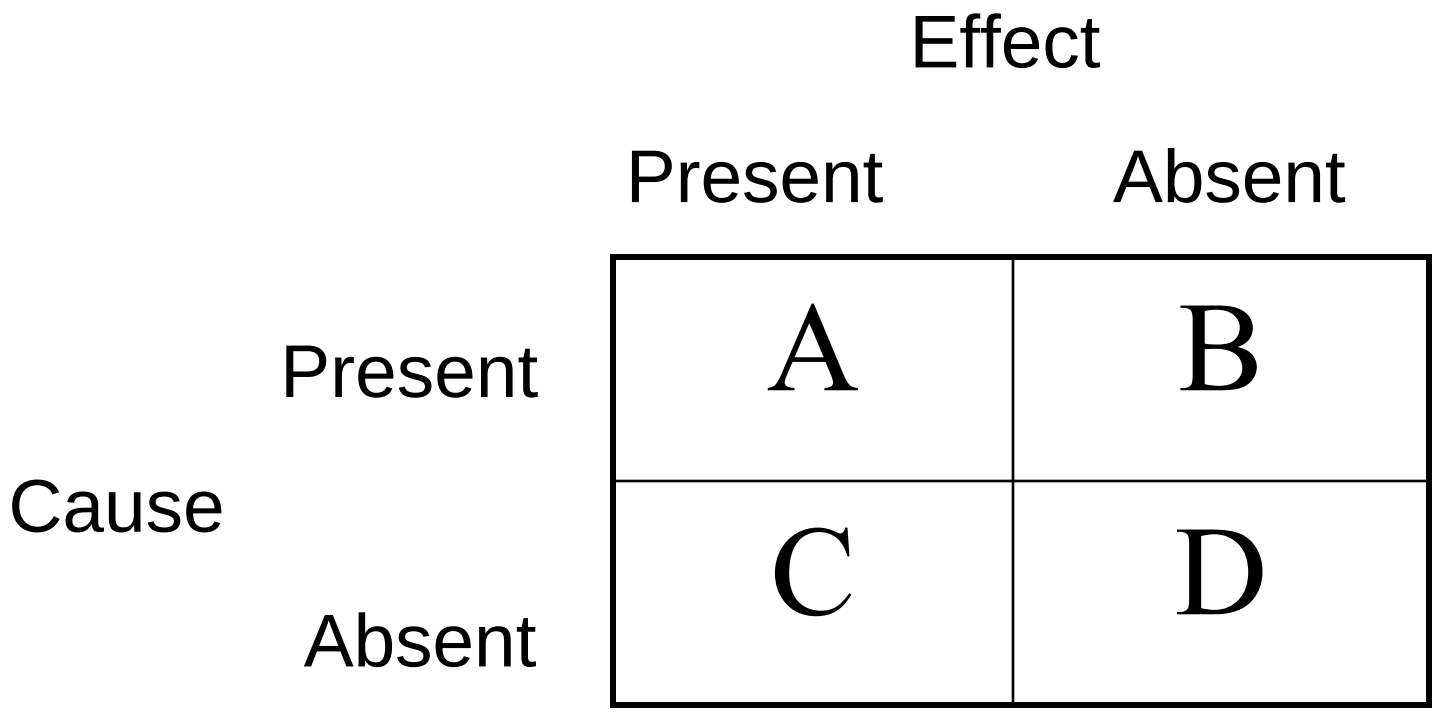
Amanda M. Kelley
Richard B. Anderson
Michael E. Doherty
Bowling Green State University

Abstract

There is limited research showing that people can draw correlational and causal inferences from data that are correlationally indeterminate, i.e., data that contain insufficient information to support computation of a correlation coefficient. In the present studies, participants made judgments about the causal relationship between two variables under conditions in which the stimulus data were correlationally indeterminate (i.e., there was zero variance in one of the variables), or determinate (i.e., the stimulus correlations were computable). Participants' judgments reflected the mathematical characteristics of the stimuli, even in the indeterminate conditions. The findings extend the range of situatuons in which indeterminate data give rise to causal inferences, and are interpreted with respect to focal set theory, representativeness, and Bayesian inference.

Introduction

- Correlation detection is an essential part of human cognition
  - Identification of causal relationships
  - Category formation
- People are sensitive to objective correlation in estimation tasks (e.g., Boynton, et al., 1997)
- Contingency tables
  - Two binary variables resulting in four possible observations
    - Cell A - cause-present effect-present
    - Cell B - cause-present effect-absent
    - Cell C - cause-absent effect-present
    - Cell D - cause-absent effect-absent


Effect
Present
Absent
Cause
Present
Absent
A
B
C
D

- People tend to make unequal use of the data in each cell (e.g., Levin, Wasserman, & Kao, 1993)

  Cell A > Cell B ≥ Cell C > Cell D

- Strategies using subset of data

  Cell A strategy, A vs. B strategy

  Under differing conditions, people make use of these simplifying strategies (Arkes & Harkness, 1983)

- Rational Models
  - In order for correlation to be computed, both factors must have variance
- Two models not requiring defined correlation
  - Rescorla-Wagner (1972) model - process model
  - Causal Support theory - Bayesian approach (Griffiths & Tenenbaum, 2005)

- Indeterminacy – data in which one variable does not vary resulting in an undefined correlational relationship
- Very few studies have looked at how people perceive indeterminate data (Clement, et al., 2002; Griffiths & Tenenbaum, 2005; White, 2000)
  - Typically find that people are willing to draw inferences from indeterminate data
  - Have not looked at this type of sample exclusively

Experiment 1

- Research Question and Rationale
  - How often will participants judge it possible to draw inferences from indeterminate samples?
  - When they do judge it possible, what kinds of judgments of causality will be made?
- Method
  - 34 undergraduate participants
  - Participants were shown data describing two variables that could be either “present” or “absent”
    - Causal candidate: chemical
    - Effect: plant growth
  - After viewing the data, participants made a judgment about the relationship between the two variables

- 5 trial types
  - Positively correlated trials (S+)
  - Negatively correlated trials (S-)
  - Zero correlation trials (N)
  - Indeterminate Present trials (Condition A is always present) (Ip)
  - Indeterminate Absent trials (Condition A is always absent) (Ia)
- Participants saw 90 trials (18 of each type)
  - 9 observations per trial

Example of an Indeterminate Present trial:

Chemical X of 90

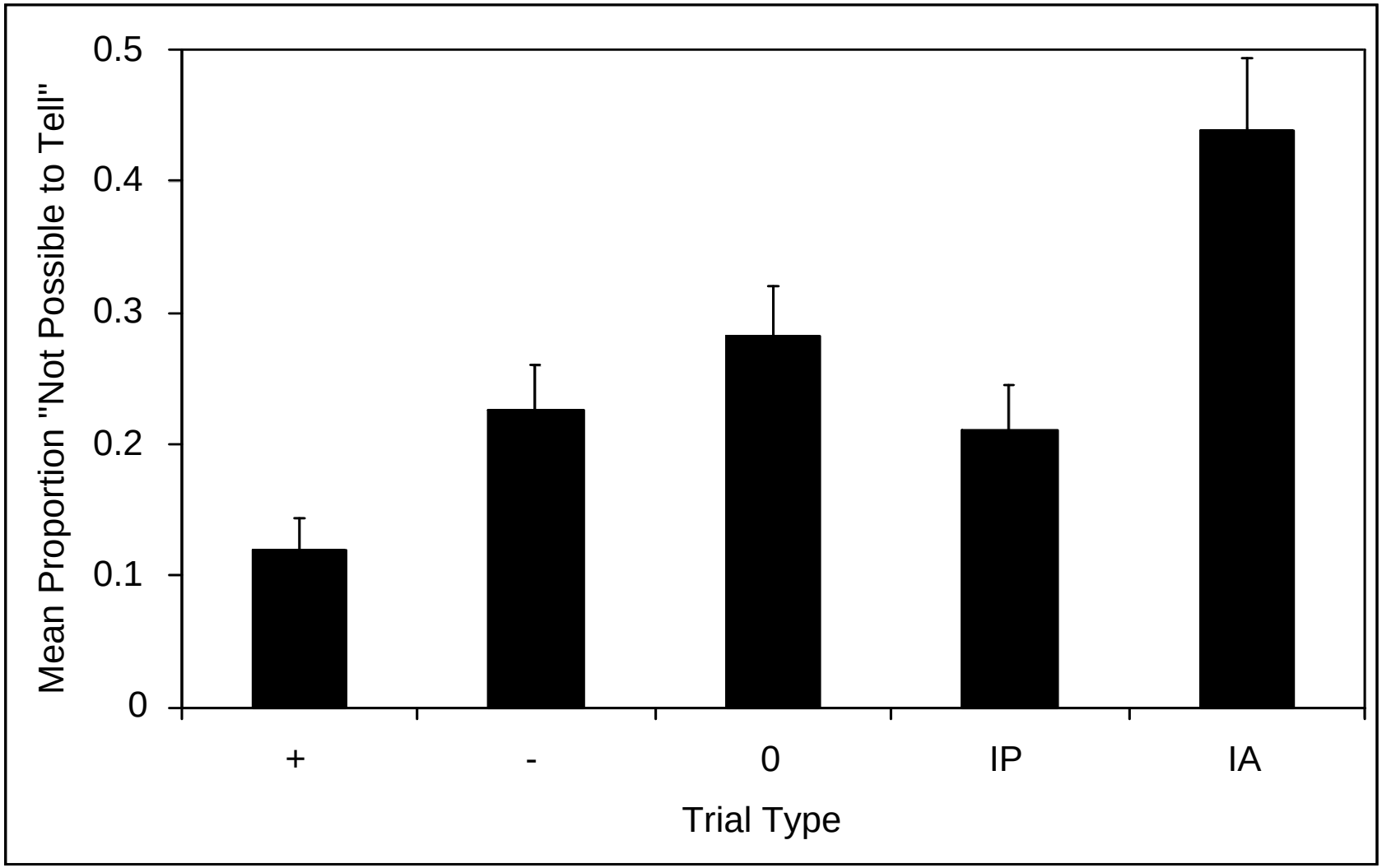
Plant #	Chemical	Plant Growth
1	Yes	Yes
2	Yes	Yes
3	Yes	No
4	Yes	Yes
5	Yes	Yes
6	Yes	No
7	Yes	Yes
8	Yes	Yes
9	Yes	Yes

Forced Choice options

1. Chemical tends to make plant grow.
   - Coded as +1 (relationship rating)
2. Chemical doesn't affect plant growth.
   - Coded as 0 (relationship rating)
3. Chemical tends to prevent plants from growing.
   - Coded as -1 (relationship rating)
4. It is not possible to tell whether there is or is not a relationship between the presence or absence of the chemical and the presence or absence of plant growth.

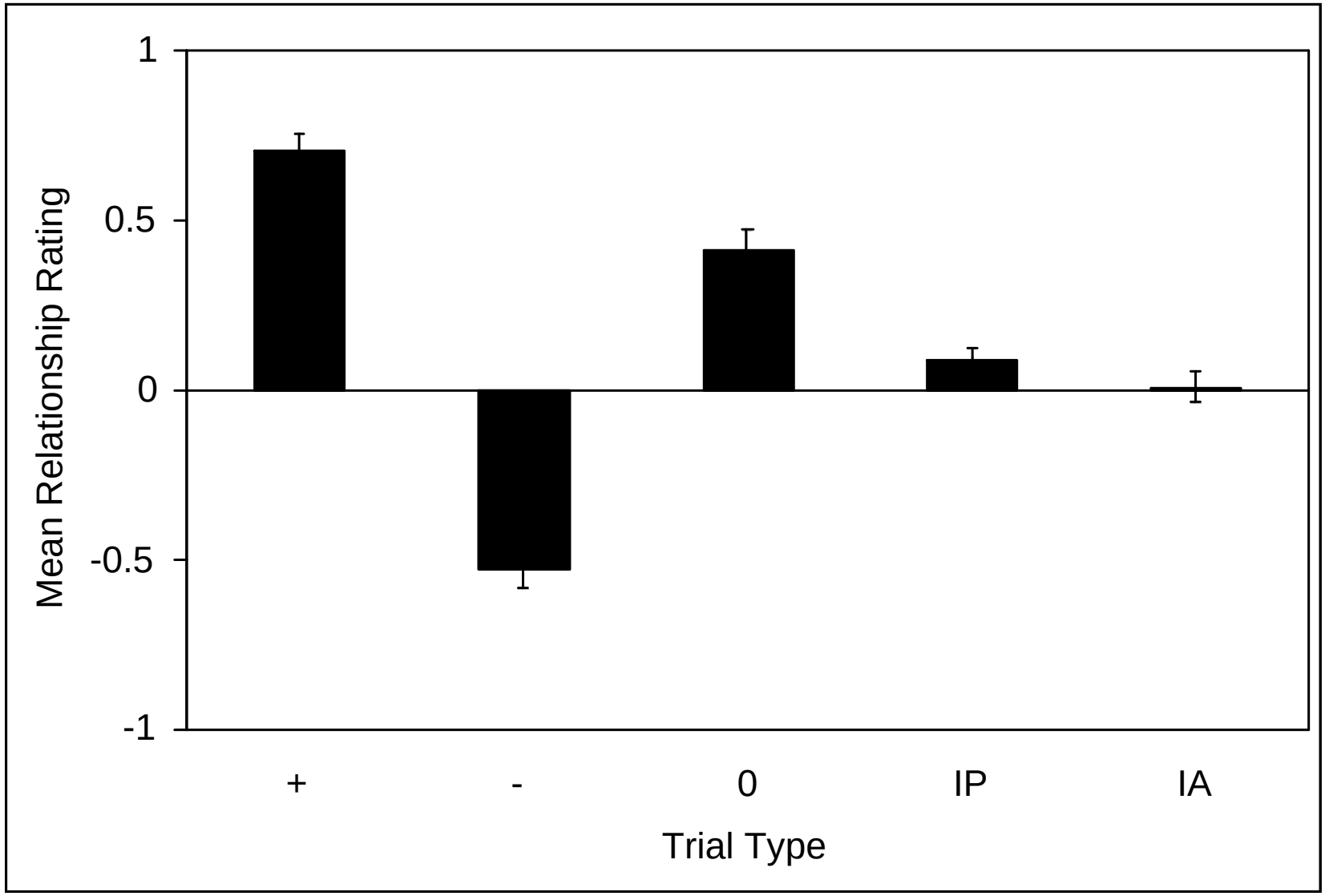
Results

- Within-subject ANOVAs with planned comparisons
- "Not possible to tell" Judgments
  - Mean proportion of "not possible to tell" judgments varied significantly across trial types, $F(1, 4) = 11.379$, $p<.001$
  - Highest mean proportion in Indeterminate Absent trials
  - Lowest mean proportion in Positively Correlated trials
  - Indeterminate present trials less similar to indeterminate absent trials than determinate trials



Note. Error bars represent the standard error of the mean. Trial types included Positively Correlated (+), Negatively Correlated (-), Uncorrelated (0), Indeterminate Present (IP), and Indeterminate Absent (IA).

- Relationship Ratings
  - Mean relationship ratings varied significantly across trial types, $F(1, 4) = 70.324$, $p<.001$
  - Indeterminate trials not significantly different from zero
    - Stimuli varied with respect to frequency ratio of A:B or C:D
    - Pattern of means shows that as A:B ratio become more extreme so do ratings, little effect for C:D ratio
  - Positive bias in zero correlation trials
  - Positively correlated trials rated as having a positive relationship
  - Negatively correlated trials rated as having a negative relationship



Note. Error bars represent the standard error of the mean. Trial types included Positively Correlated (+), Negatively Correlated (-), Uncorrelated (0), Indeterminate Present (IP), and Indeterminate Absent (IA).

Discussion

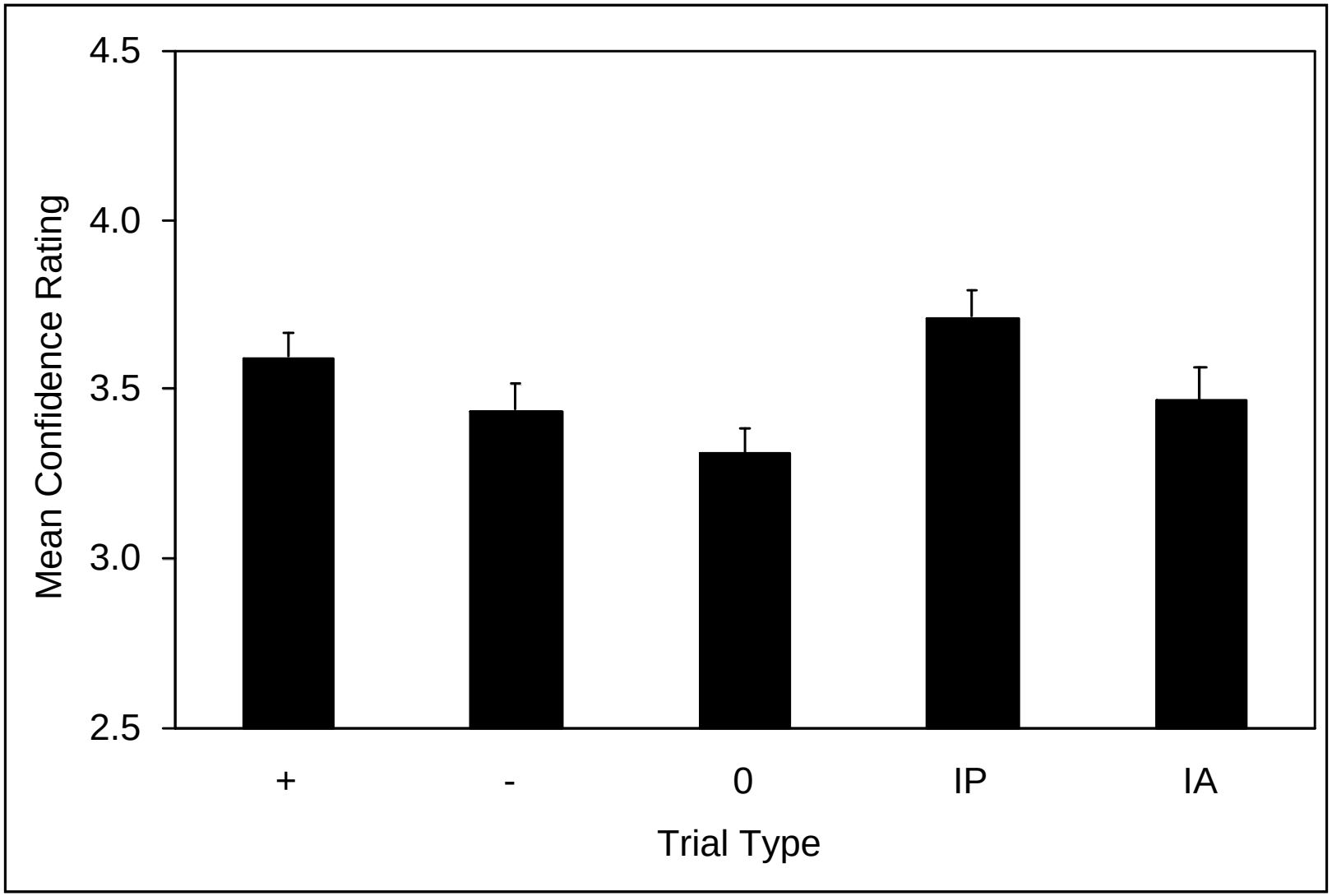
- Results for determinate trials consistent with participants computing sample correlation
  - This explanation is not applicable to the indeterminate trials
- Inconsistent mean proportion of “not possible to tell” judgments in indeterminate trials
  - Participants were not discriminating between determinate and indeterminate

Experiment 2

- Research Rationale
  - Remove possibility of missing data in indeterminate samples in order to investigate cell ratio effects on causal judgments
- Method
  - 50 undergraduate participants
  - Procedure was identical to that of Exp. 1 EXCEPT:
    - Exclusion of "not possible to tell"
    - Inclusion of confidence rating scale
      - Very low (1) to Very high (5)

Results - data analyzed at two levels

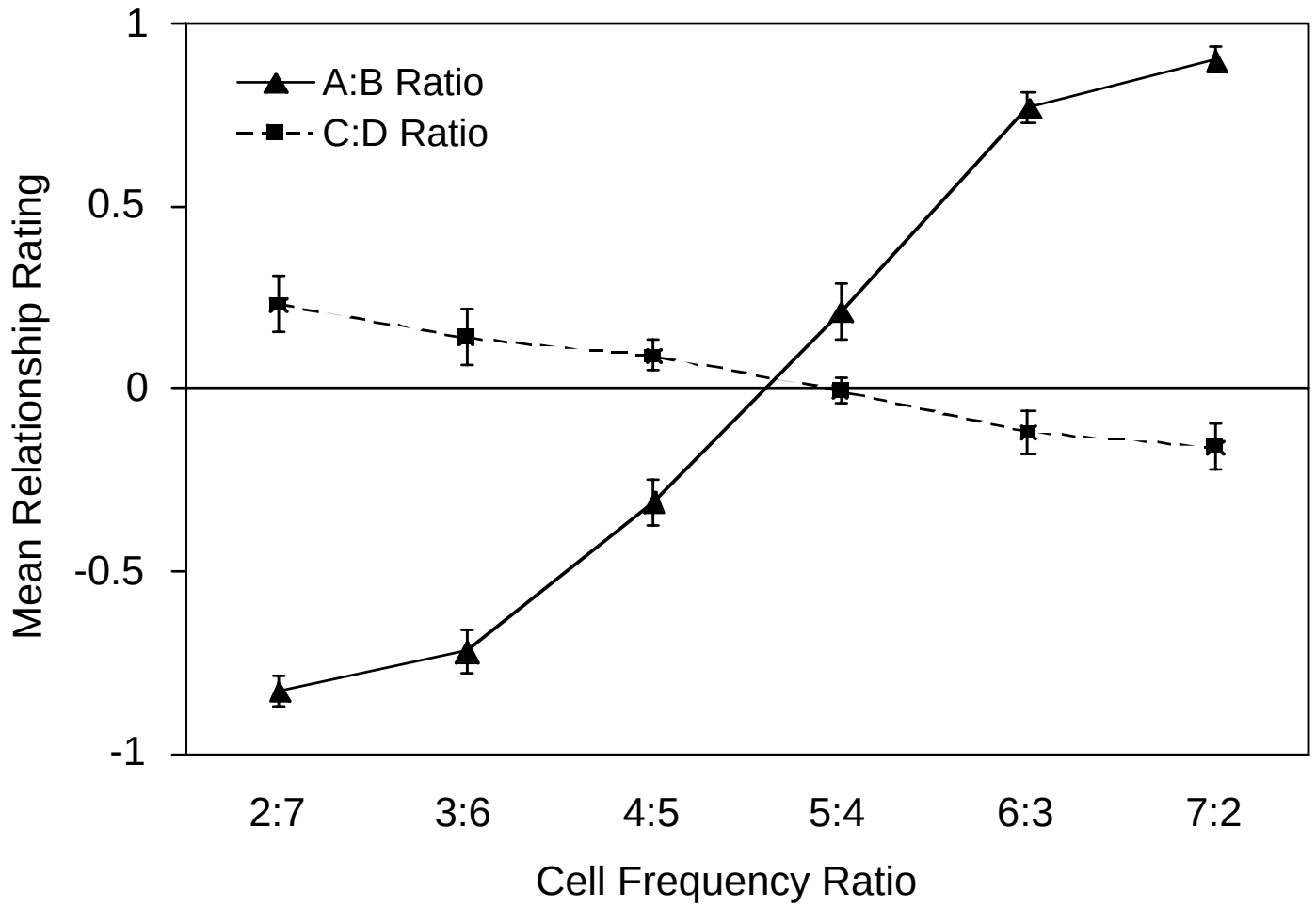
- Effects of trial type - Relationship Ratings
  - Mean relationship rating varied significantly across trial types, $F(1, 4) = 141.79$, $p < .001$
    - Replicated results of Experiment 1
  - Mean confidence ratings varied significantly across trial types, $F(1, 4) = 11.99$, $p < .001$
    - Greatest mean confidence in positive correlation and indeterminate present trials
    - Lowest mean confidence in zero correlation trials



Note. Error bars represent the standard error of the mean. Trial types included Positively Correlated (+), Negatively Correlated (-), Uncorrelated (0), Indeterminate Present (IP), and Indeterminate Absent (IA).

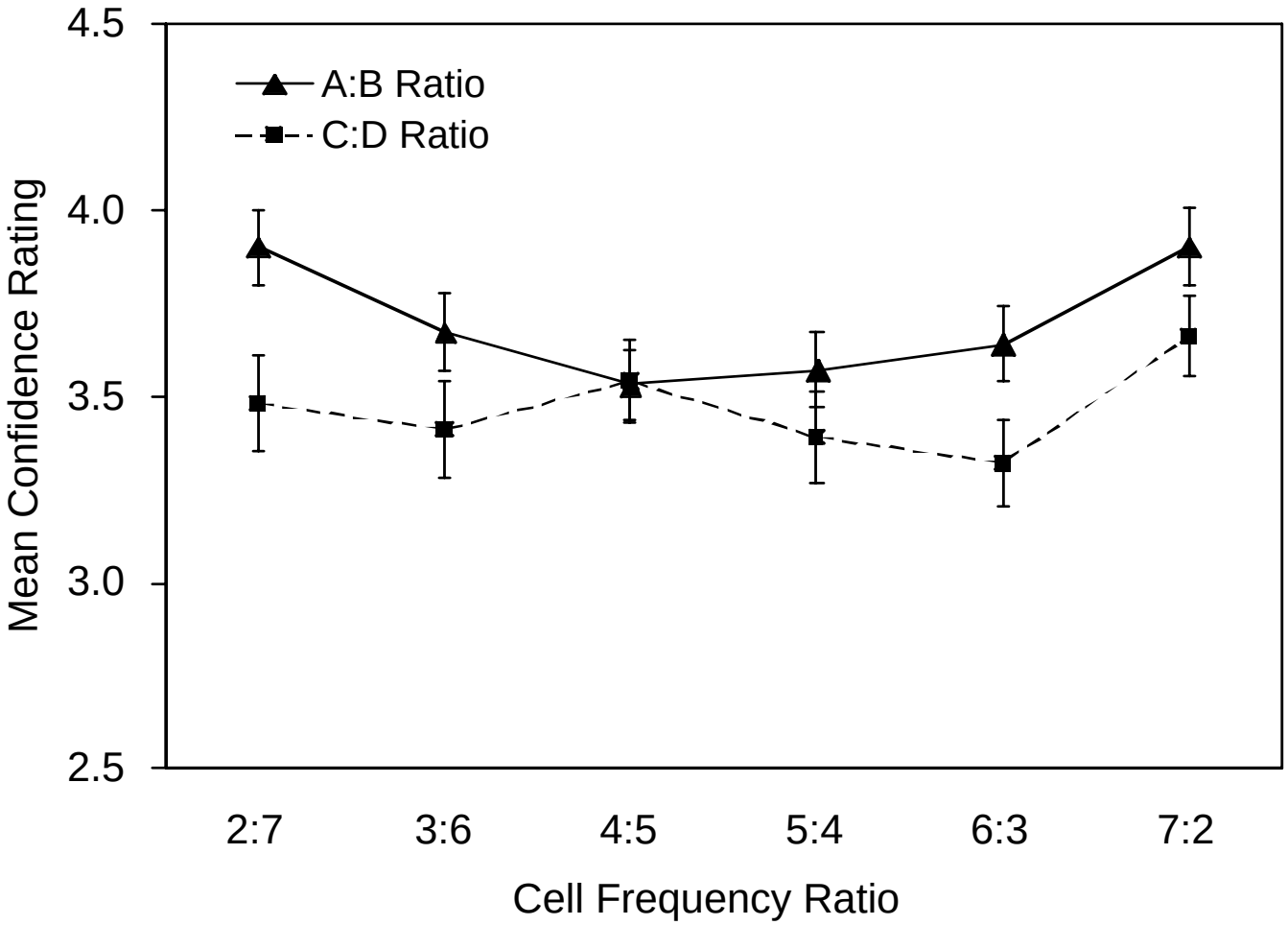
Effects of Cell Ratio (Indeterminate trials)

- Relationship Ratings
  - 6 (ratio) x 2 (indeterminate-present vs. indeterminate-absent) ANOVA
  - Main effect of ratio $F(5, 245) = 60.53$, $p < .001$
  - Significant interaction $F(5, 256) = 111.24$, $p < .001$
  - Significant linear effects of cell ratio in both indeterminate conditions



Note. Error bars represent the standard error of the mean.

- Confidence Ratings
  - Same analysis as for relationship ratings
  - Main effect of ratio, $F(5, 245) = 5.31$, $p < .001$
  - Main effect of indeterminate conditions, $F(1, 49) = 11.39$, $p < .001$
  - Interaction between the two variables, $F(5, 256) = 2.45$, $p = .034$



Note. Error bars represent the standard error of the mean.

Discussion

- Effect of trial type on relationship ratings replicated from Exp. 1
- Replicated pattern of effect of cell ratio
  - **Strong** effect of A:B ratio
  - **Small** effect of C:D ratio
  - Suggests that cause-absent information is not completely excluded and is consistent with cell preference research (e.g., Levin, et al., 1993)

GENERAL DISCUSSION

- 2 main findings
  - Perception of indeterminate trials is dependent on which cells contain observations
  - Sensitivity to cell ratio is shown in both types of indeterminate conditions
    - Suggests that all 4 cells influence judgments (however, unequally)
- Bayesian approach to understanding inferences
  - McKenzie's (in press) rarity hypothesis - emphasis on cause-present effect-present observations because they are rare in the environment hence more informative

- Griffiths & Tenenbaum's (2005) causal support theory
  - Very generally suggests that people use likelihood that sample was drawn from a variety of populations
  - Does not account for assymetry of effects in indeterminate-present vs. indeterminate-absent condtions
- Present results conform to **some** aspects of Bayesian inference but not all
- Suggests present models of causal judgments are limited in their applicability

References

Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General, 112,* 117-135.

Boynton, D. M., Smith, L. D, & Stubbs, D. A. (1997). Sensitivity and bias in covariation detection: A direct approach to a tangled issue. *Organizational Behavior and Human Decision Processes, 72,* 79-98.

Clement, M., Mercier, P., & Pasto, L. (2002). Sample size, confidence, and contingency judgment. *Canadian Journal of Experimental Psychology, 56,* 128-137.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51,* 334-384.

Kelley, A. M., Anderson, R. B., & Doherty, M. E. (under review). The perception of causation in correlationally indeterminate data. *Memory & Cognition.*

Levin, I. P., Wasserman, E. A., & Kao, S. F. (1993). Multiple methods for examining the contributions of specific cell information to contingency judgments. *Organizational Behavior and Human Decision Processes, 55,* 228-250.

McKenzie, C. R. M. (in press). A Bayesian view of covariation assessment. *Cognitive Psychology.*

Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian learning: Variations in the effectiveness of reinforcement and non-reinforcement. In A.H. Black & W.F. Prokasy (Eds.), *Classical Conditioning II: Current Theory and Research* (pp. 64-99). New York: Appleton-Century-Crofts.

White, P.A. (2000). Causal judgment from contingency information: the interpretation of factors common to all instances. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1083-1102.