Coding Effect Sizes

8 April 2020

Modern Research Methods

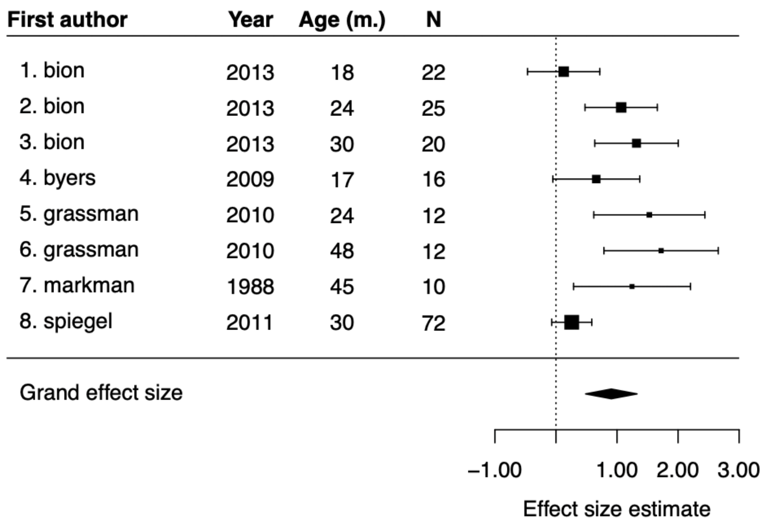
Assignment 7

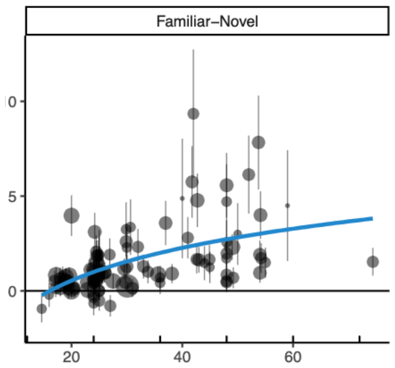
- Over 750 papers entered as a class!
- Spreadsheets due tomorrow at noon (no need to turn anything in to Canvas)

- Each of the three groups should have paper entered for the first 50 pages of google results (unless we decided on a different search protocols)
- You should have entries for *all* the studies defined by your search protocol
- For all the papers returned in your search protocol, all columns should have values
- All of your data should be in *one* spreadsheet (“relevant_studies”).
- Your coding labels should either be ‘include’ or ‘exclude’ (all ? should be resolved prior to tomorrow by talking to your group).
- See Assignment 7 sheet for details

Conducting a Meta-analysis

Final product





1. Identify Topic
2. Conduct literature search
3. Code studies and calculate ES
4. Plot and analyze data
5. Report and discuss results

Literature Search (Assignment 7)

Literature Search for the Linda Problem

File Edit View Insert Format Data Tools Add-ons Help Last edit was made yesterday at 2:00 PM by Funglun Chan

	A	B	C	D	E	F	G	H
1	protocol_id	coder_name	date_added	google_scholar_page	unique_id	paper_citation_apa	link	screening_decision
2	1	zoe	4/1/2020	1	morier1984	Morier, D. M., & Borgida,	https://journals.sag	exclude
3	1	zoe	4/1/2020	1	charness2009	Morier, Dean M., and Eu	https://www.scienc	include
4	1	zoe	4/1/2020	1	sides2002	Sides, A., Osherson, D.,	https://link.springe	include
5	1	zoe	4/1/2020	1	hertwig1999	Hertwig, R. and Gigeren:	https://onlinelibrary	include
6	1	zoe	4/1/2020	1	fiedler1988	Fiedler, K. The depender	https://link.springe	include
7	1	zoe	4/1/2020	1	bonini2004	Bonini, N., Tentori, K. an	https://onlinelibrary	include
8	1	zoe	4/1/2020	1	wolford1990	Wolford, G., Taylor, H.A.	https://link.springe	include
9	1	zoe	4/1/2020	1	agnoli1989	Agnoli, Franca, and Davi	https://www.scienc	include
10	1	zoe	4/1/2020	1	moro2008	Moro, R. On the nature c	https://link.springe	exclude
11	1	zoe	4/1/2020	2	dulany1991	Don E. Dulany and Denis	https://guilfordjourr	include
12	1	zoe	4/1/2020	2	hillel1991	Bar-Hillel, M. Commenta	https://link.springe	exclude
13	1	zoe	4/1/2020	2	fisk2016	Pohl, R. F. (Ed.). (2017).	https://www.taylorf	exclude
14	1	zoe	4/3/2020	2	reyna1991	Reyna, Valerie F. "Class	https://www.scienc	exclude
15	1	zoe	4/6/2020	2	nilson2008	Nilsson, H. (2008), Ex	https://onlinelibrary	include

Get papers (Assignment 8)

Participants were assigned to one of two counterbalanced language conditions: Language 1A and Language 1B. Eighteen additional infants were tested and excluded for the following reasons: fussiness (14), experimental error (3), and not paying attention (1). Two additional infants showed looking time preferences > 3 *SD* from the mean (one in each language group with preferences in opposite directions), and were excluded from the analyses.

Apparatus and stimulus materials—Four Italian words with a strong–weak stress pattern were selected for use in this study: *fuga, melo, pane*, and *tema* (see Table 1). Although these words were phonetically legal in English, the passages in which they were presented contained non-English phonetic features (e.g., a trill, a voiced alveolar affricate, and a palatal nasal).

We created two counterbalanced languages to control for arbitrary listening preferences at test. Language 1A consisted of three identical blocks of 12 grammatically correct and semantically meaningful standard Italian sentences (see the Appendix for sentence lists). These sentences contained the words *fuga* and *melo*, which both occurred six times in each block of 12 sentences. The component syllables of *fuga* and *melo* never appeared without each other (i.e., *fu* never appeared in the absence of *ga*, and vice versa).

Recall that the TP of, for example, *fuga* corresponds to:

$$\mathrm{TP}(ga|fu) = \frac{f(fuga)}{f(fu)}.$$

Because *fu* never appeared without *ga*, the internal TP of *fuga* (and of *melo)* was 1.0. Two other words, *pane* and *tema*, and their component syllables, were never presented in the Language 1A familiarization passages (TP = 0). In the counterbalanced Language 1B, *pane* and *tema* each occurred each six times per block (TP = 1.0), while *fuga* and *melo* (and their component syllables) never occurred (TP = 0). This design is thus exactly analogous to the original Jusczyk and Aslin (1995) study.

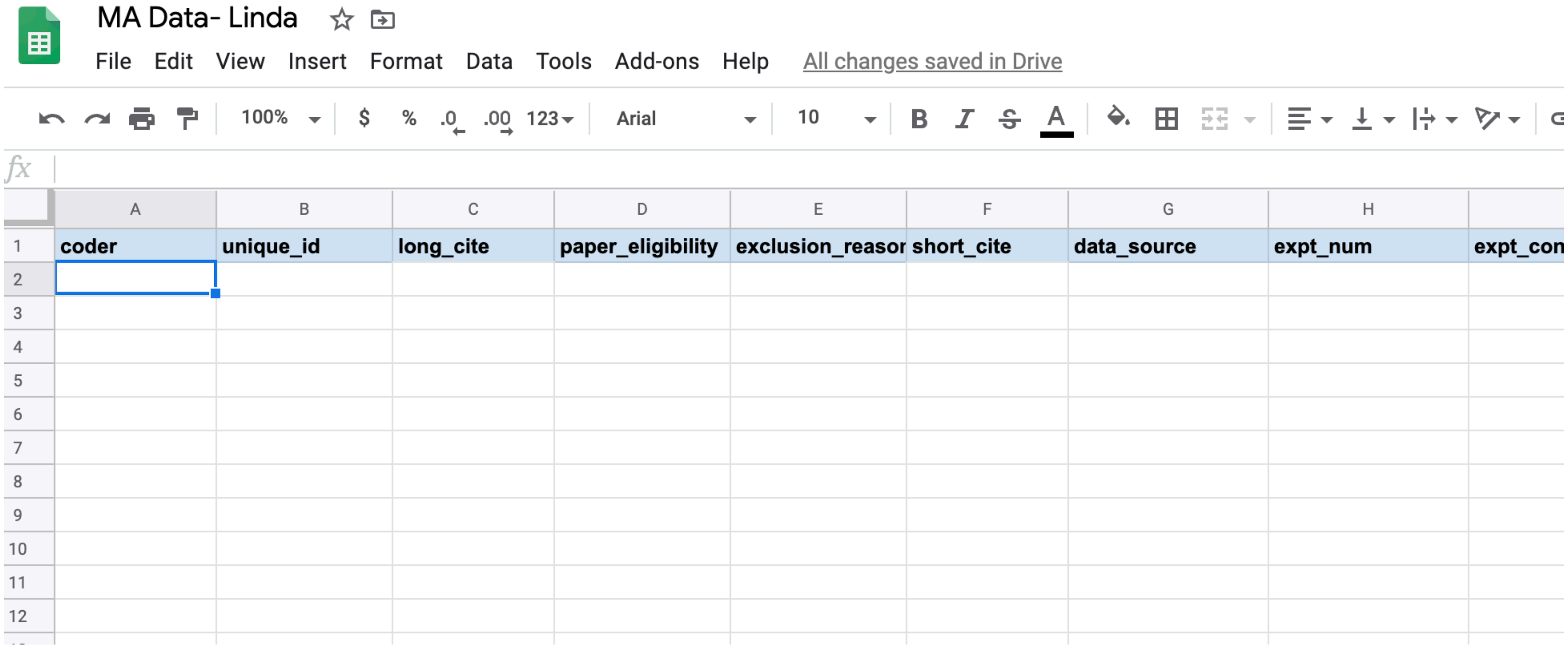
Enter data to calculate effect size (Assignment 8)

MA Data- Linda

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

	A	B	C	D	E	F	G	H	
1	coder	unique_id	long_cite	paper_eligibility	exclusion_reasor	short_cite	data_source	expt_num	expt_cor
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									

MA data



- Each group has a MA_data spreadsheet
- In lab on Friday, when your literature searches are complete, we’ll make a list of papers to code in the MA data spreadsheet

Steps for coding a paper

1. Download the paper in pdf form.
    - Try link in spreadsheet
    - In most cases, you'll need to search for the paper from the CMU library
    - https://www.library.cmu.edu/

ARTICLE / multiple sources exist. see all

On the reality of the conjunction fallacy.(Abstract)

Sides, Ashley ; Osherson, Daniel ; Bonini, Nicolao ; Viale, Riccardo

Memory & Cognition, March, 2002, Vol.30(2), p.191(8)

PEER REVIEWED

Available Online >

TOP

VIEW ONLINE

CITATIONS

LINKS

DETAILS

SEND TO

Citations

Find sources citing this or sources cited in this

View (55) citations in **SciVerse Scopus** >

Clarivate Analytics View (51) citations in **WEB OF SCIENCE** >

Links

Report a Problem >

Signed in and still can't find what you're looking for? Let us know

Ask a Librarian >

View Online

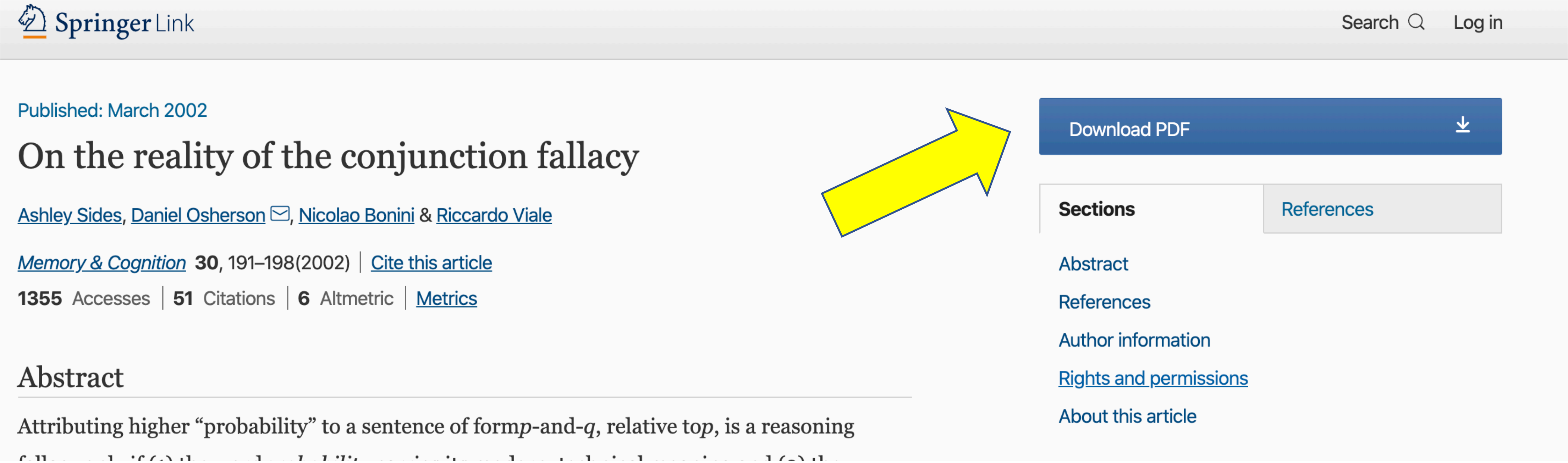
Full text availability

North East Research Libraries (NERL) Springer Journals

Available from 1997 volume: 25 issue: 1.

SpringerLink Journals - AutoHoldings

Available from 01/01/1997 volume: 25 issue: 1.


Springer Link

Search Log in

Published: March 2002

On the reality of the conjunction fallacy

Ashley Sides, Daniel Osherson, Nicolao Bonini & Riccardo Viale

Memory & Cognition **30**, 191–198(2002) | Cite this article

1355 Accesses | **51** Citations | **6** Altmetric | Metrics

Abstract

Attributing higher “probability” to a sentence of form*p*-and-*q*, relative to*p*, is a reasoning

Download PDF

Sections References

Abstract

References

Author information

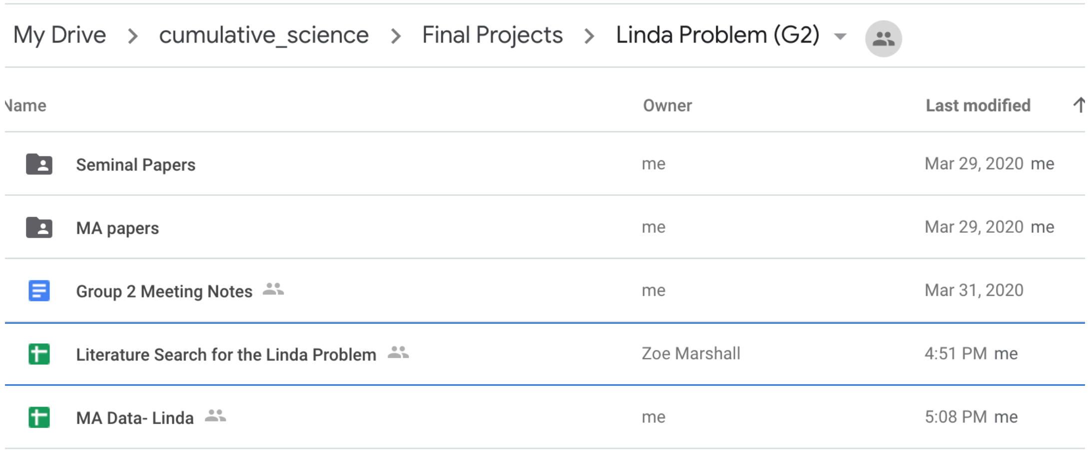
Rights and permissions

About this article

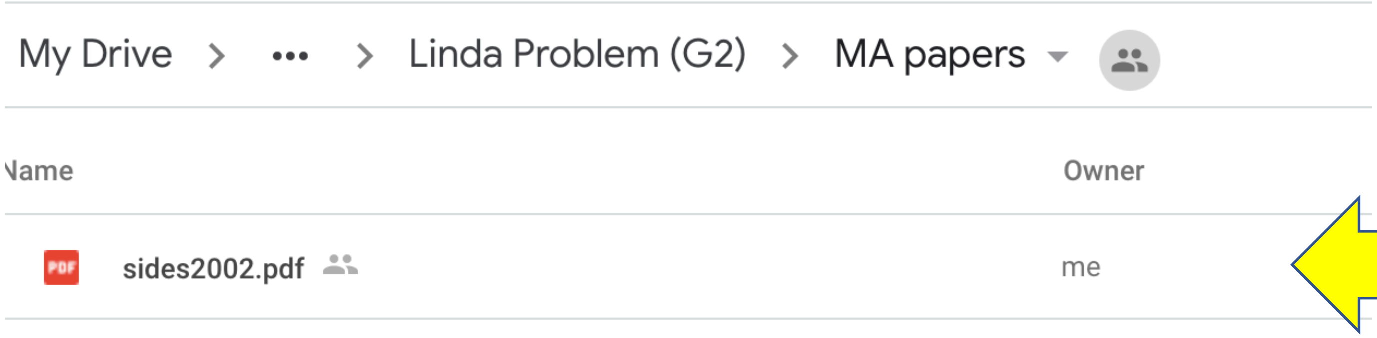
- Some papers you may not be able to get access to
- If you can’t get access to the full pdf of a paper, make a note in the MA_data spreadsheet
- If you’re having trouble getting access to *most* papers, let us know.

Steps for coding a paper

2. Put it in the folder titled `MA_papers`



3. Relabel the paper with the `unique_id`



Steps for coding a paper

4. Skim the paper to determine whether it satisfies your inclusion criteria.

- If it does not, fill in the relevant information in your `MA data` spreadsheet, and move on to the next paper.

On the reality of the conjunction fallacy

ASHLEY SIDES, DANIEL OSHERSON, NICOLAO BONINI, and RICCARDO VIALE
Rice University, Houston, Texas

Attributing higher "probability" to a sentence of form *p*-and-*q*, relative to *p*, is a reasoning fallacy only if (1) the word *probability* carries its modern, technical meaning and (2) the sentence *p* is interpreted as a conjunct of the conjunction *p*-and-*q*. Legitimate doubts arise about both conditions in classic demonstrations of the conjunction fallacy. We used betting paradigms and unambiguously conjunctive statements to reduce these sources of ambiguity about conjunctive reasoning. Despite the precautions, conjunction fallacies were as frequent under betting instructions as under standard probability instructions.

The Conjunction Fallacy

Here is the famous Linda story, to be labeled *E* (for *evidence*) in what follows:

(*E*) Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations.

The task is to rank various statements "by their probability," including these two:

(B) Linda is a bank teller.
($B \wedge F$) Linda is a bank teller and is active in the feminist movement.

A majority of respondents across a variety of studies ranked $B \wedge F$ as more probable than B (see Hertwig & Chase, 1998, for a review of findings; the original report is Tversky & Kahneman, 1983). This judgment is in apparent violation of the conjunction law $\Pr(X \wedge Y \mid Z) \le \Pr(X \mid Z)$ for any statements *X, Y, Z*, with strict inequality for nontrivial cases such as the present example.

The law is not violated, however, if participants in these studies understand the word *probability* in a sense different from the one assigned to it by modern probability theory. There is similarly no violation if *B* is interpreted to mean $B \wedge \neg F$ or is interpreted in any way other than as a conjunct of $B \wedge F$. The need for clarity about these issues is discussed in the remainder of the present section. We then describe experiments in which we attempted to provide a sharper test of the thesis that naive conjunctive reasoning can be led into fallacy.

Let us first note that we do not attempt to defend naive reasoning by denying the defective character of the judgment $\Pr(X \wedge Y \mid Z) > \Pr(X \mid Z)$ (if such a judgment is ever made). In particular, we believe the concept of probability can be sensibly applied to single events (like man reaching Mars before 2050) and is governed by principles familiar from discussions of Bayesianism (as in Earman, 1992; Horwich, 1982; and Howson & Urbach, 1993). All the events that figured in our experiments were singular in character, resisting placement in classes of similar cases that allow for a meaningful frequency count.

Interpreting the Word Probability

As documented in Hertwig and Gigerenzer (1999), *probability* is polysemous in the general population. It has often been noted, moreover, that, through much of its premodern history, the term *probable* carried a connotation of "approvable opinion" (see Hacking, 1975, chap. 3). Appeal to authority was one way that an opinion was approvable, but another was via evidential support. Thus, John Locke (1671) defined probable propositions as those "for which there be arguments or proofs to make it pass or be received for true" (cited in Krause & Clark, 1993, p. 71). A respondent working with the latter interpretation of probability would attempt to determine whether *E* provides more support for *B* or for $B \wedge F$. In what follows, we formalize support in a familiar way and observe that it justifies the intuition that *E* provides greater support for $B \wedge F$ than for *B*. Several alternative formalizations would serve our purposes just as well, but we do not attempt a survey of possibilities. Our point is that at least one plausible reading of *probable* exculpates reasoners from the conjunction fallacy.

Many authors agree that a statement *X* supports a statement *Y* to the extent that $\Pr(Y \mid X)$ exceeds $\Pr(Y)$ (see the references cited in Fitelson, 1999, in which the term *confirmation* is used in place of *support*). A simple way to quantify this relation is via the quotient $\Pr(Y \mid X) / \Pr(Y)$ (the difference works just as well). Here Pr denotes probability in the modern, technical sense, and the quotient $\Pr(Y \mid X) / \Pr(Y)$ translates the support concept into modern terms. According to the definition, *E* supports $B \wedge F$ more than *E* supports *B* if and only if

$$\frac{\Pr(B \wedge F \mid E)}{\Pr(B \wedge F)} > \frac{\Pr(B \mid E)}{\Pr(B)}.$$

The authors thank Andrea Cerroni, Karin Dudziak, Denise Wu, Andrea Pozzali, and Zhihua Tang for assistance in performing Experiment 1. Correspondence should be addressed to D. Osherson, Department of Psychology, MS-25, P. O. Box 1892, Rice University, Houston, TX 77251-1892 (e-mail: osherson@rice.edu).

Steps for coding a paper

5. Find the relevant data in the paper for coding the effect size, and highlight it using your pdf reader.

- If the pdf is too old, then draw a red line next to the relevant information. This is just so we can find the information later if we want to check it.

Table 1
Two Examples of Each Kind of Event Pairs Used in Experiments 1 and 2

X	$Y \wedge X$
1. The percentage of adolescent smokers in Texas will decrease at least 15% from current levels by September 1, 1999.	The cigarette tax in Texas will increase by $1.00 per pack and the percentage of adolescent smokers in Texas will decrease at least 15% from current levels by September 1, 1999.
2. By September 1, 1999, an experimental vaccine for childhood leukemia will be announced.	The National Institutes of Health (NIH) will increase spending on vaccine development by 50% in the first 9 months of 1999, and by September 1, 1999, an experimental vaccine for childhood leukemia will be announced.
X	$Y \wedge Z$
3. The University of Houston Philosophy Department will hire 3 new faculty members by September 1, 1999.	The University of Houston writing department will be rated in the top 10% nationwide and will announce that it will expand its faculty by September 1, 1999.
4. By September 1, 1999, Texas will require people to pass a literacy test before serving on a jury.	By September 1, 1999, Texas will start selecting juries from a pool of licensed drivers rather than registered voters and the number of registered voters will increase by 10%.
X	Y
5. Bill Clinton will announce his intention to seek a divorce before September 1, 1999.	By September 1, 1999, Janet Reno will announce her intention to run for the Presidency.
6. Fidel Castro will be removed from political power in Cuba by September 1, 1999.	U.S. forces will be sent to Havana, Cuba before September 1, 1999.

Out of 9 possible occasions, the average number of errors per participant was 3.4 ($SD = 2.55$). In the betting condition, 36 of 45 participants committed at least one conjunction error (choosing to bet on a conjunction rather than its conjunct), with an average of 3.2 errors ($SD = 2.33$). A t test revealed that the difference in means did not approach significance. Table 2 shows the number of participants in each condition who made m errors, for m between 0 and 9.

For a given fallacy item and a given condition, call the proportion of participants who committed the conjunction error the *fallacy score* for that item in that condition. We correlated the fallacy scores for the 9 items across the two conditions. The Pearson coefficient was .82 ($p < .01$), suggesting similar mental processes in the two conditions.

No fallacy was committed, of course, if the participants' responses resulted from inattention or lack of interest. Inattentive responding would favor equal fallacy rates across the nine different fallacy items. But, in fact, the 9 items attracted very different numbers of fallacy responses in both the betting condition and the probability condition. For example, combining across the two conditions, Item 1 in Table 1 attracted 51 fallacy responses (out of 90 possible), whereas Item 2 attracted 19 fallacy responses. Thirty-four participants committed the fallacy for Item 1 but not for Item 2, whereas only 2 participants had the reverse profile. This difference was reliable by a binomial test in which inattentive responding was assimilated to the toss of a fair coin ($p < .05$, two-tailed). The same test yielded a reliable difference between Item 1 and every other item except for two. We conclude that our participants' fallacious responding was not due to inattention to the task.

EXPERIMENT 2

To test the robustness of our findings, Experiment 1 was replicated with a new group of students at a different university.

Method

Stimuli. With a few exceptions, the stimuli from Experiment 1 were used for Experiment 2. The exceptions arose from events transpiring during the interval separating the experiments. For example, the event "the new *Star Wars* movie will receive two thumbs down from Siskel and Ebert by September 1, 1999" needed replacement after the death of Gene Siskel early that year.

Participants. The participants in Experiment 2 were 57 undergraduate volunteers from the University of Houston, a public institution with diverse enrollment, located in downtown Houston.

Procedure. The procedure was the same as described above for Experiment 1. Twenty-nine participants were randomly assigned to the probability condition, and 28 participants were randomly assigned to the betting condition. The participants were run in groups of 3–14 in the first months of 1999.

Results

In the probability condition, all 29 participants committed at least one conjunction error, with an average of 5.93

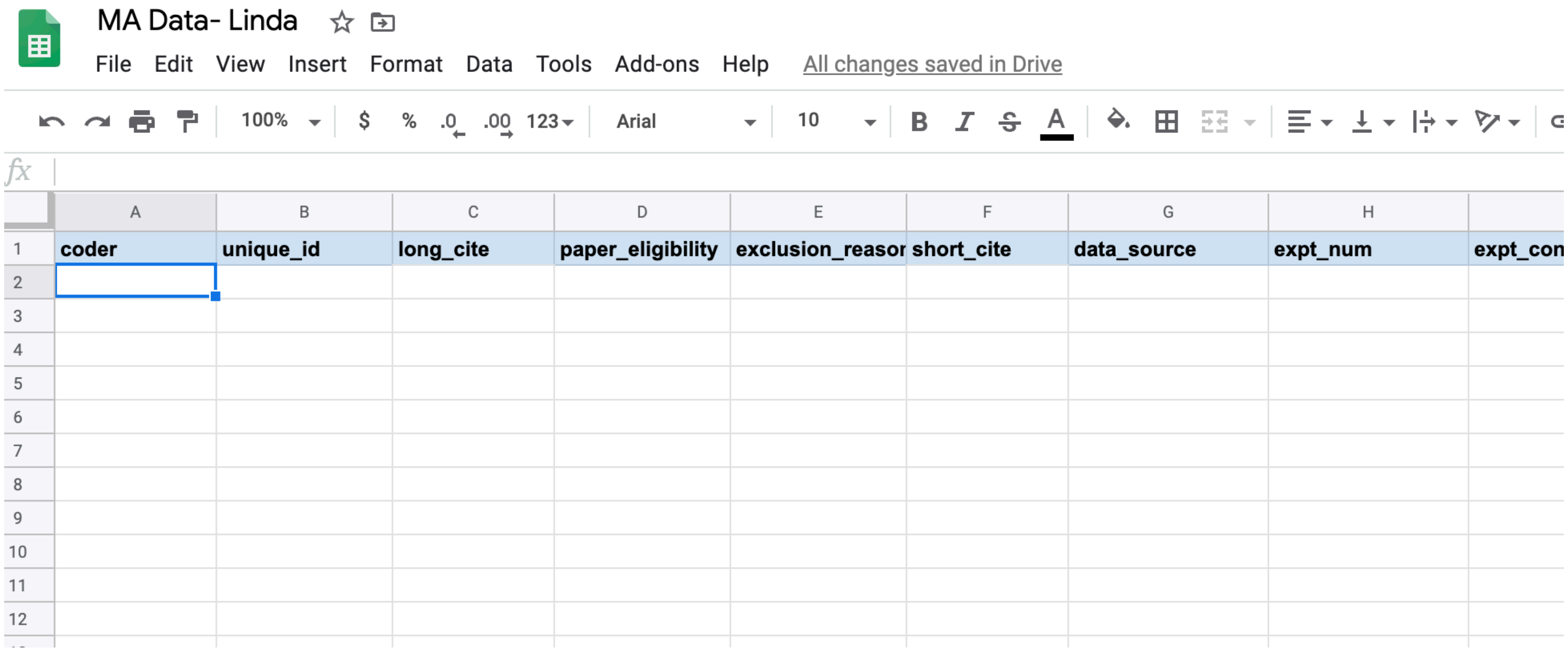
Table 2
Number of Participants in the Probability and Betting Conditions of Experiment 1 Who Committed m Conjunction Errors ($0 \leq m \leq 9$)

	Number of Conjunction Errors									
Condition	0	1	2	3	4	5	6	7	8	9
Probability	7	8	1	6	8	5	4	3	2	1
Betting	9	5	3	6	7	8	5	1	0	1

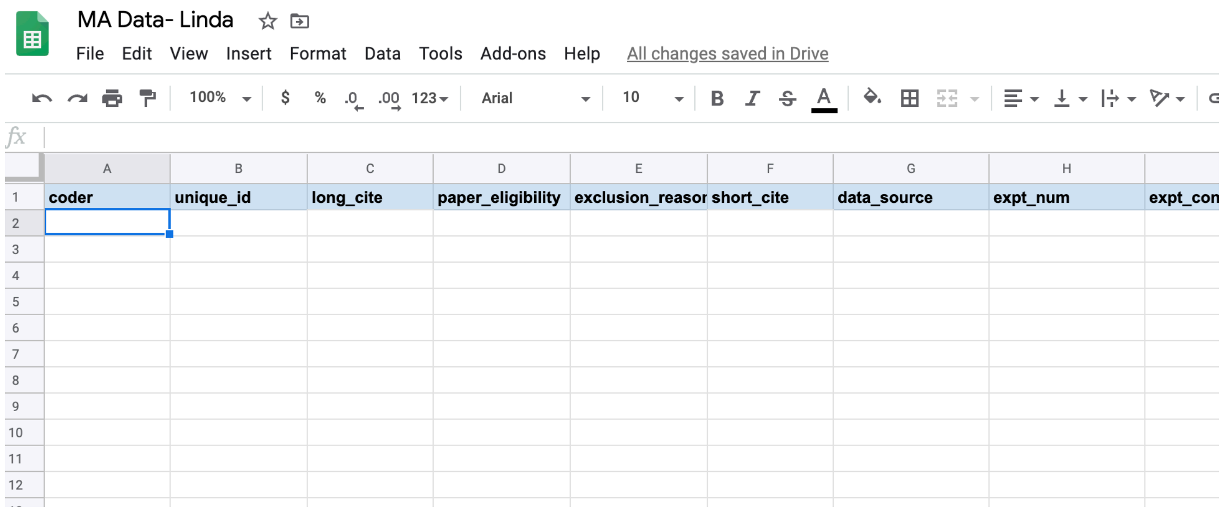
Note—For each condition, $n = 45$ participants.

Steps for coding a paper

6. Enter the data in your `MA data` spreadsheet



Coding studies for effect sizes



studies

Meta-info about paper

	A	B
1	study_ID	long_cite
2	SaffranAslinNewport1996	Saffran, J. R., Aslin, R. N., & Nev
3	SaffranAslinNewport1996	Saffran, J. R., Aslin, R. N., & Nev
4	PelucchiHaySaffran2009a	Pelucchi, B., Hay, J. F., & Saffran
5	PelucchiHaySaffran2009a	Pelucchi, B., Hay, J. F., & Saffran
6	PelucchiHaySaffran2009a	Pelucchi, B., Hay, J. F., & Saffran
7	PellucchiHaySaffran2009b	Pelucchi, B., Hay, J. F., & Saffran

...

Stats for calculating effect size

AP	AQ
x_1	SD_1
7.97	2.008581589
6.77	2.155550974
8.21	2.593838854
9.08	2.95160973
8.75	2.03646753
10.06	NA

...

moderators

AJ	AK
method	exposure_phase
HPP	familiarization
HPP	familiarization
HPP	familiarization
HPP	familiarization
HPP	familiarization
HPP	familiarization

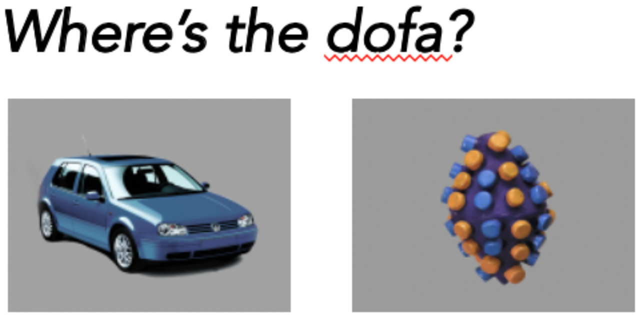
⋮

MA data template

- Data are tidy: each row is a single observation
- In our case, observation = effect size
- Each paper may have multiple effect sizes in it.
- In some cases, may not report the means and SDs you need to calculate an effect size, but there are other ways to get the ES – we'll talk about this for each group separately
- Most of the papers that you plan to include won't end up having usable data in it.
- MA data spreasheet varies by MA, but broadly the same for everyone

Effect Sizes

- Quantifies "success" in effect in each paradigm



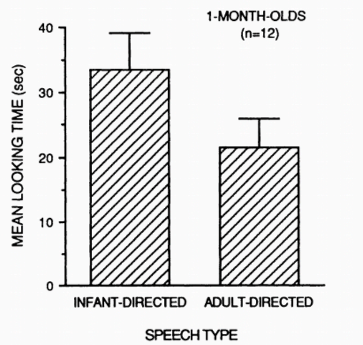


FIG. 2.—Mean looking times (in sec) of 1-month-old subjects from Experiment 1 (including standard errors); ID = infant-directed and AD = adult-directed.

- What does success mean for each of the three phenomena we're looking at?

Minimal Group Paradigm

- ingroup liking/allocation vs. outgroup liking/allocation
- Remember dependent measure has to be either liking or allocation
- Convert months to age (1 month = 30.44 days)
- To calculate effect size, we need one of the follow:
  - $M_{ingroup}$, $M_{outgroup}$, $SD_{ingroup}$, $SD_{outgroup}$
  - *t*-score comparing in group vs. outgroup
  - Cohen's *d*
- If there's a plot the the means and standard deviation on them (error bars) we can measure the plot – let me know if you encounter one of these.
- Moderators to code

Linda Problem

- Linda problem: Mean number of conjunctive errors AND number of participants that made at least one error
- To calculate effect size, we need one of the follow:
  - Prop. with at least one error
  - M_{errors}, $SD_{errrors}$
  - *t*-score comparing in N errors to chance (I bet you won’t see this)
  - Cohen’s *d* (I bet you won’t see this)
- If there’s a plot the the means and standard deviation on them (error bars) we can measure the plot – let me know if you encounter one of these.
- Moderators to code

Syntactic Bootstrapping

- Convert months to age (1 month = 30.44 days)
- Moderators
- Just verbs?
- Include screenshot of video stimuli