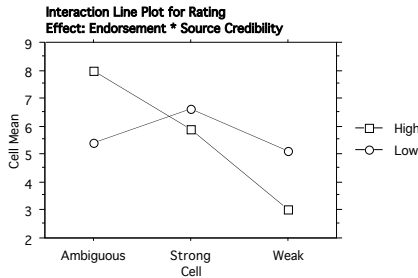


1a. How seriously do people take product reviews? Chaiken and Maheswaran (1992) conducted an interesting experiment in which they varied the credibility of the review source and the general message of the review. They asked college students to read a review of a new telephone answering machine. The researchers told half the participants that the review came from a flyer printed by the discount store Kmart (low credibility) or from the magazine Consumer Reports (high credibility). Each participant then read one of three types of review, an unambiguous strong review, an ambiguous review (the answering machine was better than some machines but not as good as others), or an unambiguous weak review. The researchers then asked the participants to rate on a 10-point scale their willingness to buy the answering machine for \$50 (10 = very willing to buy). The results of their study are replicated below. Complete the analysis and interpret the results of this study as completely as you can (as in a Discussion section). [15 pts] [Pittenger]

ANOVA Table for Rating

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Source Credibility	1	.067	.067	.056	.8143	.056	.056
Endorsement	2	80.433	40.217	33.618	<.0001	67.235	1.000
Source Credibility * Endorsement	2	58.233	29.117	24.339	<.0001	48.678	1.000
Residual	54	64.600	1.196				



Means Table for Rating
Effect: Source Credibility * Endorsement

	Count	Mean	Std. Dev.	Std. Err.
High, Ambiguous	10	8.000	1.155	.365
High, Strong	10	5.900	1.101	.348
High, Weak	10	3.000	1.155	.365
Low, Ambiguous	10	5.400	.699	.221
Low, Strong	10	6.600	1.174	.371
Low, Weak	10	5.100	1.197	.379

There is a significant main effect of Endorsement ($F(2,54) = 33.618, p < .001$) and a significant interaction between Credibility and Endorsement, $F(2,54) = 24.339, p < .001$. Thus, you should focus your attention on the interpretation of the interaction.

$$HSD = 4.19 \sqrt{\frac{1.196}{10}} = 1.45$$

So, either:

For Low Credibility Source, Strong reviews lead to significantly higher willingness to buy than Weak reviews. However, for High Credibility Source, Ambiguous reviews lead to higher willingness to buy than Strong reviews, which led to higher willingness to buy than Weak reviews.

or

For Weak reviews, Low credibility sources lead to greater willingness to buy than High credibility sources. However, for Ambiguous reviews, High credibility sources lead to greater willingness to buy than Low credibility sources. For Strong reviews, there is no difference between Low and High credibility sources.

1b. The F-Value (and Power) are quite low for Source Credibility. Would you then conclude, given the lack of a main effect for Source Credibility, that Source Credibility had no impact on the dependent variable (willingness to buy)? Why? [3 pts]

No. There is an effect of Source Credibility, but it shows up in the interaction term rather than in the main effect.

1c. Suppose that these same 60 data points had been analyzed in a single-factor ANOVA on Endorsement. What would the source table look like for that single-factor ANOVA? [7 pts]

Source	df	SS	MS	F
Endorsement	2	80.43	40.21	18.62
Error	57	122.9	2.16	
Total	59	203.33		

2. How many participants would you need for these designs with a minimum $n = 30$? [10 pts]

A completely independent groups (between) 3x5 design	450
A completely independent groups (between) 2x9 design	540
A completely repeated measures (within) 2x9 design	36
A mixed 3x7 design, with the second factor repeated measures (within)	126
A mixed 2x9 design, with the first factor repeated measures (within)	270
A mixed 2x9 design, with the second factor repeated measures (within)	72

3. False memories emerge when people study a list of words, all of which are related to a specific “lure” word. For instance, you might be presented with a list of words such as: garbage, waste, can, refuse, sewage, bag, junk, rubbish, sweep, scraps, pile, dump, landfill, debris, and litter. Although the word “trash” never appeared in the list, people will often report having seen “trash.” This paradigm was developed by Deese and later revived by Roediger and McDermott, so it is often referred to as the DRM paradigm. Dewhurst and Robinson (2004) were interested in the extent to which such false memories (or intrusions) vary with age. To that end, they had 5-, 8-, and 11-year olds study lists of words (such as the one above) anticipating a memory test. Their errors were examined (as one dependent variable) to see if they varied with age. The type of error a child might make was classified as Phonological (e.g., a rhyming intrusion, such as “bunk” instead of “junk”), Semantic (e.g., reporting “trash” even though it never occurred, but was semantically related to the list), or Unrelated (e.g., reporting that they had seen “shovel”). Dewhurst and Robinson got results such as those seen below. On the basis of these results, tell me what you would expect them to find in their analysis of the data. Then tell me what sense you would make of these data, as in a discussion section. [20 pts]

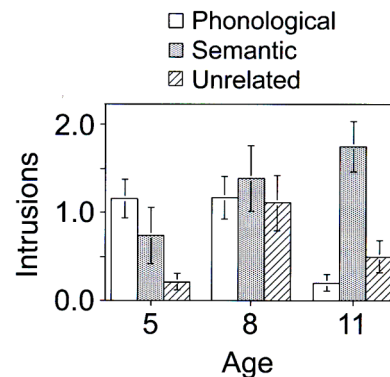


Fig. 2. Mean numbers of phonological, semantic, and unrelated intrusions per child (with standard errors) as a function of age group.

AGE: There appears to be a main effect of Age, with 5-Year-Olds ($M = \sim.7$) and 11-Year-Olds ($M = \sim.8$) producing fewer intrusions than 8-Year-Olds ($M = \sim1.1$).

ERROR TYPE: There appears to be a main effect of Error Type, with Semantic Errors ($M = \sim1.3$) occurring more frequently than Phonological Errors ($M = \sim.9$) or Unrelated Errors ($M = \sim.5$).

AGE x ERROR: For 5-Year-Olds, Phonological Intrusions occur more frequently than Semantic or Unrelated Intrusions. The situation is different for 8-Year-Olds, for whom the type of intrusion didn't appear to differ that much. Finally, the situation differs again for 11-Year-Olds, with Semantic Intrusions dominating over the Phonological and Unrelated Intrusions.

Language usage appears to change as children mature. When 5-year-olds make errors on this task, they are rarely unrelated to the list, nor are they semantic intrusions. Instead, they seem to be phonological, which may well dominate the way these children have words organized (on the basis of sound). Eight-year-olds show an increase in semantic errors, which probably is a function of increasing semantic arrangement of information in their heads, but they also still make phonological errors at a level consistent with the five-year-olds. I'm not sure why their unrelated intrusions increase. Finally, note that the eleven-year-olds exhibit strong semantic organization, as their intrusions are largely semantic (as is true of adults on this task), with very few phonological or unrelated intrusions.

4. We have discussed both experimenter effects and participant effects.
- Describe the Rosenthal study that showed the problems that might emerge from the experimenter learning about how the participants were performing on a memory experiment *or* the Rosenthal study that showed the impact of telling the experimenter what to expect in terms of the ratings of pictures of people's faces.
 - What antidote(s) would you suggest to deal with such expectancy effects?
 - Define demand characteristics and provide a clear example. [10 pts]

Respond to this question based on your notes regarding the social psychology of the experiment.

5. Dr. Harold Hedd was interested in the impact of men's hair on women's ratings of the men's attractiveness. He was also interested in the extent to which women's age might have an impact on their ratings. He took thirty pictures of men with full heads of hair and then used computer software to give the men in the pictures a receding hairline and also to make them completely bald. He then presented thirty pictures to women and asked them to rate the attractiveness of the men in the pictures on a 7-pt scale (1 = unattractive and 7 = attractive). One-third of the pictures showed the men with a full head of hair, one-third of the pictures showed the men with receding hairlines, and one-third of the pictures showed the men as bald. To ensure that the effects were not specific to particular men, the men's pictures were rotated through all three options, such that some women saw a man with a full head of hair, other women saw that man with a receding hairline, and other women saw that man as bald-headed. The women in the study were young (median age of 26), middle-aged (median age of 48), or older (median age of 62). Complete the source table and interpret the results as completely as you can. [10 pts]

ANOVA Table for Rating

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Age	2	6.578	3.289	3.289	.0488	6.578	.581
Hair	2	121.378	60.689	60.689	<.0001	121.378	1.000
Age * Hair	4	.356	.089	.089	.9854	.356	.066
Residual	36	36.000	1.000				

Means Table for Rating

Effect: Age * Hair

	Count	Mean	Std. Dev.	Std. Err.
Middle-Age, Bald	5	5.600	1.140	.510
Middle-Age, Full Hair	5	6.000	1.000	.447
Middle-Age, Receding	5	2.200	.837	.374
Older, Bald	5	5.000	1.225	.548
Older, Full Hair	5	5.600	1.140	.510
Older, Receding	5	2.000	1.000	.447
Young, Bald	5	6.200	.837	.374
Young, Full Hair	5	6.400	.894	.400
Young, Receding	5	2.800	.837	.374

There is a main effect for Age and a main effect for Hair, but no interaction. To determine the nature of the effects, you need to compute HSD:

$$HSD = 3.45 \sqrt{\frac{1.0}{15}} = .89$$

Thus, for Age: Young men receive significantly higher ratings ($M = 5.1$) than Old men ($M = 4.2$), with Middle-Age men ($M = 4.6$) falling in between and not differing significantly from either.

For Hair: Full hair ($M = 6.0$) and bald men ($M = 5.6$) don't differ but both lead to significantly higher ratings than for men with Receding hairlines ($M = 2.3$).