|  |
| --- |
| “A meta-analysis of the effects of demand characteristics” - could it just be “A meta-analysis of demand characteristics”? or “demand characteristic effects”? |

I had originally considered similar titles and had some minor hesitations. But I’m realizing now that my concerns were perhaps pedantic. (E.g., I felt that you can only meta-analyze *effects*, not constructs or theories. But plenty of titles in *Psychological Bulletin* say things like “Meta-analysis of theory X”.)

For now, I changed the title to “A meta-analysis on demand characteristics”. I’ll note, though, that APA style suggests that we should also summarize the key finding in the title. In that case, the title would be: “A meta-analysis on demand characteristics: Demand characteristics generally produce small, but highly heterogeneous, acquiescence effects”. I have a small preference for the former, though.

|  |
| --- |
| “If this scenario were real, you would reasonably question whether Colesology is a valid method of scientific inquiry…”  I think I worry that this intro might be a little too cute for the paper even though it works well as a talk. you could say the same thing but more directly. “Demand characteristics, the idea that experimental participants react to the their beliefs about the desired result of an experiment, are a surprisingly mysterious part of experimental methods in psychology. They are a methodological artifact that can be invoked in almost any situation, and neither their direction nor their magnitude is known in advance.” |

I had the same concern, but I wanted to get your impression first.

I like your proposed re-write and have updated the manuscript accordingly. For now, I tried to limit the amount of re-writing and restructuring required—but I’m open to starting from scratch if it’s not coming together nicely.

|  |
| --- |
| “In 1962, Martin Orne published a seminal paper highlighting a view that challenged deeply-ingrained beliefs about experimental psychology.”  I tend to agree with folks who counsel psych writing to be idea focused rather than person-focused, so e.g. using parenthetical citation to attribute. “From a current perspective, it may seem obvious that research participants actively try to make sense of the experimental experience. Yet this idea was highly controversial when it was first proposed (Orne, 1962).” |

I’m happy to try that writing style here. I’m sure you noticed that “active person-centric” writing style was beat into my brain over the years.

|  |
| --- |
| You might spell out the methodological progression in these tests of DCs over the years, with Mummolo and Peterson (2019) getting a couple of lines or even a para. |

That’s an interesting idea. I have so much to say about the methodological history that I was planning on saving it for another paper. I’m oversimplifying things, but I think it’s useful to think about three different methodological goals:

1. Documenting demand characteristics as a methodological concern (early work by Orne in the 50’s and 60’s)
2. Understanding how demand characteristics operate (big focus in 60’s and 70’s, which largely culminated and ended with the Rosnow and Rosenthal model).
3. Ruling out concerns about demand characteristics (60’s onward)

To *document* demand characteristics as a methodological concern, Orne and colleagues introduced several methodological advancements. The *quasi-control method*, where would-be-participants review the protocol and comment on demand characteristics (Orne 1969, 1970). If they identify places where there are demand characteristics, then the artifact is a concern. *Quasi-experiments,* *non-experiments*, or *as-if experiments* where participants are generally asked to respond “as-if” they’ve received the experimental manipulation. This method raised serious concerns about research on hypnosis because they found that the behavior of participants who faked hypnosis was indistinguishable from those who received an actual hypnotic induction (Orne, 1959).

To understand how demand characteristics *operate*, researchers typically have manipulated demand characteristics. However, the manipulation is not always explicit; some researchers manipulate instructions or a cue they assume conveys their hypothesis (e.g., Edward, 2002); some broadly state the goals of the study without providing a clear directional prediction (Irwin, 1968); some explicitly and clearly state a hypothesis (Sigall et al. 1970). Sometimes, demand characteristics are manipulated in a context that provides information about mechanism. For example, Sigall et al. (1970) (a) told participants a hypothesis, but then (b) added that confirming the hypothesis would be evidence of a mental issue. Participants didn’t conform to the stated hypothesis, so they argued that participants use information about the purpose of the study to bolster self-presentation—not help the experimenter. To the best of my knowledge, researchers only recently began measuring variables that may provide information about mechanism (Coles et al., 2022; Lush et al., 2021; current meta-analysis).

The above methods have also been used to try to *eliminate* concerns about demand characteristics, which I think is the most common reason why you’ll see the phrase “demand characteristics” appear in a paper. For example, Xia et al (2014) used a demand characteristics manipulation to identify and remove demand-susceptible participants from their primary study. Many researchers use funnel debriefing in an attempt to identify and remove demand-susceptible participants. Some include individual difference measures that might measure susceptibility to demand (e.g., social desirability measures; West, Mogilner, and DeVoe, 2021). Hignett et al. (2018) had their demand characteristics additionally refer to outcomes that they did not expect to be changed by the manipulation. After not finding changes on those tangential outcomes, they concluded that demand characteristics are not problematic. Deception (Zanna et al., 1970), replications in naturalistic settings (Taylor-Covill & Eves, 2016), and difficult-to-control outcomes (Graham et al. 2020) are also used.

One reason why I avoided talking about this in depth is because I think it warrants a thorough examination of the strengths and limitations of each approach (particularly because we then focus on experimental manipulations). Furthermore, in my opinion, a lot of questionable assumptions are required for these methods to convincingly rule out concerns about demand characteristics. In fact, I’ve argued with Lush offline that claims about the impact of demand characteristics are ultimately unfalsifiable. The best we can seem to do is minimize concerns about demand.

|  |
| --- |
| could this statement of goals be more precise: “The goal of the current paper is to use meta-analysis to take stock of what we know—and what we don’t know—about this methodological artifact.” - what are the things that will emerge? direction, magnitude, consistency?  maybe that “after over 60 years” para could be split into a gap/goals paragraph and an outline/contributions paragraph? |

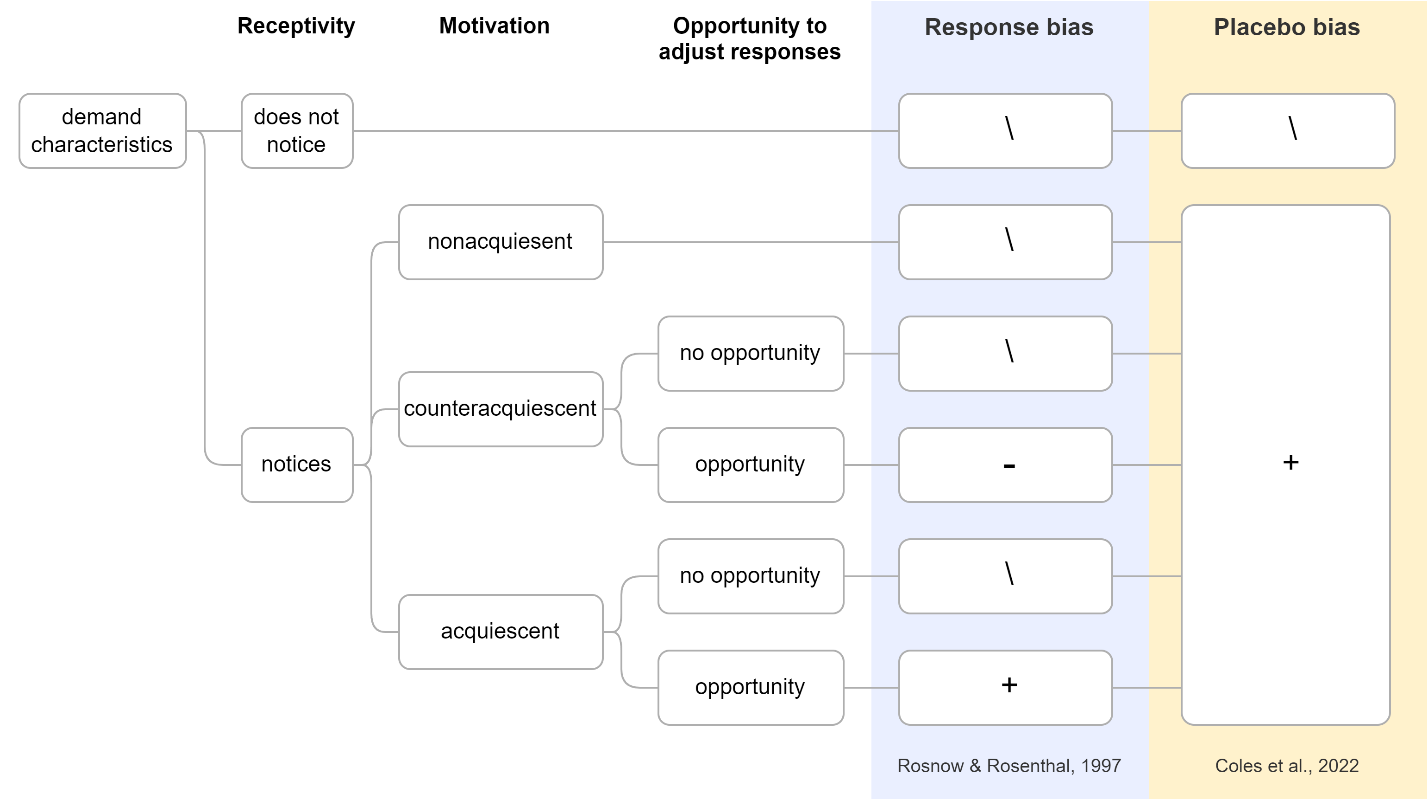
Updated the section to be more precise and to also mention how the paper is structured as multiple “studies”.

|  |
| --- |
| re the diagram. I had some suggestions for making it a little prettier/more schematic - less text on the branches of the tree and more on the labels:    the diagram suggestions connect with a broader point about potentially restructuring the exposition of how to think about demand… your frame led me to expect that the model is R&R97 vs. Cetal22. but it’s not - as far as I can tell - rather, you want those two columns to be labeled as “demand” and “placebo” and maybe you could mark them with +s and -s. (again, reducing the person-centered nature of the report). then there are two factors being posited. and they are additive (the pluses) but maybe you don’t have to have a plus (or connect the lines)? the outcomes then are just different columns that relate to particular leaves of the tree… |

I accepted most of the suggested changes to the figure, with a few minor exceptions:

1. I did not replace “demand characteristics” with “participant receives manipulation”. The reason is because I want this diagram to capture how demand characteristics would operate in scenarios where they are not explicitly manipulated. (E.g., the more typical study where demand characteristics are present, perhaps by accident or by necessity.)
2. I relabeled the blue and yellow boxes as “response bias” and “placebo bias”, but I put the citations down at the bottom. I think it is interesting and useful to point out that demand characteristic researchers have not historically considered placebo bias as a potential mechanism. To the best of my knowledge, the effect has always been conceptualized as a response bias. (Side note: Most of this work was happening in social psychology, and I think this reflects the way that social psychologists were thinking about behavior at this time.) Coles et al. (2022), Corneille and Lush (2022), and this meta-analysis demonstrate, though, that this was a mistake. Decades old advice about how to avoid the bias of demand characteristics—such as reducing motivation and opportunity—are not sufficient for reducing bias.

Updated figure:



|  |
| --- |
| So, then I wondered if there was a way to put the exposition of R&R and Coles together into one discussion of demand and placebo effects, rather than two different headings. |

I could see that working, and it would also address your concern that demand/placebo are defined too late in the paper.

The section could be structured as follows:

* There are at least two mechanisms that could underlie demand effects: responses biases and placebo biases.
* Define response bias and placebo bias.
* Discuss that demand characteristics researchers have historically been focused on response bias. Thus, they historically focused on motivation and opportunity to adjust responses (e.g., R&R). This led to practical recommendations about how to avoid the artifact (hide the purpose of the study, motivate participants to respond naturally, and/or use measures participants can’t readily adjust)
* Cetal2022, though, argued that placebo is an *overlooked* second mechanism. Thus, limiting motivation and opportunity, thus, isn’t sufficient.

|  |
| --- |
| before you launch into methods we need a paragraph or two talking about how the MA will test these hypotheses, that you’re going to code for these factors… |

Agreed. Maybe this would be in the first couple paragraphs of Study 2

|  |
| --- |
| on that note, I think you are really going to have to justify why you exclude placebo effects… we need a bit of a conceptual framework for demand as a specific kind of experimental artifact, and then when you introduce placebo, you need to tell us a bit more about why?  maybe one thing you could do is to say that this is a paper about METHODS, not about PLACEBO… so you are INVOKING placebo as a potential mechanism but MEASURING the methodological artifact. |

I would love to brainstorm this issue bit more.

In my opinion, placebo and demand characteristic research has a lot of *unrecognized* overlap. The two literatures differ in the sense that placebo research is often focused on clinical outcomes (e.g., drug response) and demand research is focused on non-clinical outcomes (e.g., conformity, perception, attitudes). Methodologically, though, they share similarities: they both seek to understand their phenomena by manipulating the stated (or implied) hypothesized effect.

One possibility is that we state that we are focusing on non-clinical experimental psychology research. A second possibility is that we can say that we are trying to focus on the demand characteristics literature territory because doing so will allow us to examine the role of placebo effects in a domain where it isn’t traditionally discussed.

|  |
| --- |
| I wonder if the positive, nil, negative demand continuum could be estimated better with network meta-analysis? https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5247317/ |

I’ve been eager to conduct a network meta-analysis for some time now. However, a key assumption in network meta-analysis is *transitivity*, which I understand as meaning that there is low heterogeneity and limited confounding. Given how high the estimated heterogeneity in this dataset is, I’m not really sure if I would trust the results of the network meta-analysis.

|  |
| --- |
| what’s the difference between nil demand and control? I think I get this but would like to have it said explicitly early on (maybe I missed it, reading in chunks) |

This was defined in the Screening section. To help clarify, I added text in that section that clarifies how the condition differs from control:

* *positive demand* (participants told that the dependent variable will increase)
* *negative demand* (participants told that the dependent variable will *decrease*)
* *nil demand* (participants told the dependent variable will be *unaffected*)
* [new text] *control* (participants not told about the relationship between an independent and dependent variable)

|  |
| --- |
| as I’m reading I’m wondering whether you would like to split the paper in to two studies - 1) the MA, clear and simple, no conceptual frameworks, and 2) the extra analysis using RR97 and the vignette study (more complicated, potentially innovative). I think this might make the paper clearer and easier to follow rather than having combined methods. |

Neat idea. Let’s try it.

Here’s how the paper might be organized:

**Introduction**

* Introduce demand characteristics as an artifact.
* Review evidence of upward biasing, downward biasing, and inconsistency across studies (e.g., Mummolo and Peterson)
* Provide overview of objectives
  + Study 1:
    - Estimates direction, magnitude, and consistency of effect.
    - Looks at potential study feature moderators (e.g., participant payment)
  + Study 2:
    - Will use ad-hoc ratings from a new set of participants to examine three mechanisms that may drive the effect: motivation, opportunity, and belief
    - Will also look at whether the new set of participants can predict the effect of demand characteristics
  + Study 3?
    - Conceptual replication of Coles et al. (2022) and additional test of the moderating role of motivation, opportunity, and belief (this is currently squeezed into the discussion)
* Study 1
  + Literature search
  + Screening
  + Effect size index
  + Types of demand characteristic comparisons
  + Overall effect, heterogeneity, forest plot
  + Study feature moderator analyses
  + Publication bias analyses
* Study 2
  + Overview of response and placebo bias
  + Explanation of post-hoc measures of potential mechanisms
  + Results from mechanism moderator analyses
  + Discussion of the limitations of the approach
* Study 3?
* Discussion: TBD

|  |
| --- |
| I think I didn’t understand why the section “Accounting for different demand comparisons.” came where it did |

This is only important for the motivation, opportunity, and belief moderators.

Based on the vignettes, participants provide these ratings for a *single* condition (e.g., how much they would be motivated to confirm the hypothesis in a positive demand condition). But because Cohen’s *d* involves a comparison between *two* conditions, the ratings from the two conditions need to be combined. This section attempts to explain how that’s done.

|  |
| --- |
| “Nonetheless, Rodgers and Pustejovsky (2021) demonstrated that the method retains fairly good statistical properties when” got repeated twice, which felt a bit awkward |

Lazy writing on my part. Fixed now :)

|  |
| --- |
| I was expecting a forest plot as the first plot … |

I’ll add one in.

|  |
| --- |
| when I saw the overall MA effect my first thought was, how does it work to put negative demands in the mix and average over them. can you say explicitly here how this estimate deals with that? |

This is stated in the effect size index section. Do you think it’s worth repeating here, though?

Here’s what I put for now:

“Results indicated that, overall, explicit manipulations of demand characteristics create a small acquiescence effect…In other words, when explicitly told a hypothesis, participants responses tend to shift in a manner consistent with that hypothesis. For example, if participants were hypothetically told that an intervention should improve mood (positive demand), they would generally report slightly improved moods; if told that an intervention should worsen mood (negative demand), they would generally report slightly worsened moods.”

|  |
| --- |
| reading the results I am more strongly in favor of my suggestion above, that is, separating the post-coded moderators from the other moderators as a second study. One reason is that there are some interesting things in the “other moderators” section! these get short shrift when you focus on the post-coded vignette moderators. |

I agree. I don’t have a *whole lot* to add to these discussions because (a) I suspect the moderators are confounded, and (b) the moderators have often been proposed in a very speculative manner (e.g., some people say students are more susceptible to demand because they care about the university, the researcher, etc.; some people say Mturkers are more susceptible because they want good ratings).

That being said, if you and I chat about the moderators a bit more, we may find that there’s a lot more to talk about in the paper!