Evolution and Human Behavior 39 (2018) 479–489

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.elsevier.com/locate/ens

A reanalysis of crossed-dimension "Who Said What?" paradigm studies, using a better error base-rate correction

David Pietraszewski

Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

ARTICLE INFO

Keywords:
Who Said What?
Social categorization
Coalitional psychology
Race
Sex
Age
Accent

ABSTRACT

Cognitive and evolutionarily-minded researchers have increasingly adopted the "Who Said What?" memory-confusion paradigm, a powerful and sensitive paradigm originating from social psychology which allows researchers to unobtrusively measure social categorization. The paradigm has been particularly important over the past two decades for arbitrating between different functionalist hypotheses about the evolved social mind. Bor (2018) has pointed out, however, that the simple arithmetic base-rate correction inherited from social psychology for this paradigm is problematic. This standard base-rate correction—in use since 1992 and in over a dozen studies—creates a mathematical artifact in which the calculated magnitude of categorization by one dimension can affect the calculated magnitude of categorization by a second, crossed dimension, even when the two dimensions are in fact fully independent from one another. No one had noticed this in 25 years. Worryingly, this means that all previously-reported "Who Said What?" studies featuring two crossed dimensions have reported potentially-biased estimates of the true magnitude of categorization. Here, a reanalysis of four large research projects is presented, involving 56 effect sizes across 31 between-subjects conditions (Pietraszewski, Cosmides, & Tooby, 2014; Pietraszewski & Schwartz, 2014; Pietraszewski, Curry, Petersen, Cosmides, & Tooby, 2015; Pietraszewski, 2016). In all cases, results from those papers hold or are even strengthened by the reanalysis. In particular, the previously-reported experimental reductions in racial categorization were in fact underestimates. This reanalysis also reveals that the most common artifact of the previously-used error correction was to artificially-inflate categorization by the weaker of the two dimensions—at least among the studies reported here. Finally, a succinct list of best practices for use of the "Who Said What?" paradigm in the future is presented, including but not limited to this new base-rate correction.

1. Introduction

One of the virtues of science is that is it collaborative in the broadest sense of the word. Multiple sets of eyes can look at the same sets of problems, paradigms, or assumptions, all over a period of years, decades, and even centuries. Occasionally, one of those sets of eyes may see something new, such as a new solution to a long-standing problem or an assumption that no one had previously noticed or thought to question. Bor (2018) has provided such a service to the increasingly-used "Who Said What?" memory confusion paradigm, which elegantly and unobtrusively measures social categorization. To understand this new insight, we must first understand the logic of the paradigm itself.

1.1. The logic of the "Who Said What?" paradigm

Precisely what the "Who Said What?" paradigm does is that it allows us to measure source-tag confusions in memory. For example, suppose Bob tells you, "The sea looks red today". A memory will be created for the content [*the sea looks red today*]. But that content will not just float around freely in your mind. Instead, information-processing *tags* will also be attached. These tags will include, for instance, information about the creedal (or truth) value of the content, and the source or origin of that content, such as who said the statement (Cosmides & Tooby, 2000; Johnson, Hashtroudi, & Lindsay, 1993). The tags that mark source of that content, *source-tags*, serve a number of critical information-processing functions. They allow otherwise contradictory information to peacefully co-exist. For example, suppose Sue tells you that instead of red, "The sea looks blue today". The source-tags [*Sue*] and [*Bob*] allow each of the two conflicting representations about what color the sea looks like to be cordoned off from one another. Source tags also provide retrieval addresses for recalling memories. For example, the question, "What did Bob tell you the other day?" would be impossible to answer if the information that he provided was not also tagged to him as its source.

E-mail address: davidpietraszewski@gmail.com.

https://doi.org/10.1016/j.evolhumbehav.2018.04.005
Received 18 January 2018; Received in revised form 20 April 2018; Accepted 30 April 2018

The "Who Said What?" paradigm takes advantage of these source-tags to measure how an event is processed by the mind. When Bob tells you the see looks Red, an entire complex suite of cognitive adaptations process the event. For example, because Bob is a conspecific, cognitive adaptations for assessing conspecific features will be activated. These will assess sex and age, formidability, attractiveness, and so on. All the while, the outputs of these processes—that Bob is [*male*], [*middle-aged*], and so on—are being generated. What is happening in memory at the same time is that these generated outputs are becoming tagged or associated with the content [*the sea looks red today*].

As researchers, we can take advantage of this source-tagging procedure to determine what features of people or of events are being processing and attended to. For example, if we want to know if the mind has processed Bob's sex, we need some way to measure if and how much that output [*male*] was generated. This requires relying on *confusions* between source-tags, rather than simply looking at the source-tags themselves. If we just ask participants, "Was "*The sea looks red today*" said by Bob?", we can be certain that some constellation of the features that Bob embodies was noticed or not. But we can never be sure about *which* features were recalled to answer the question correctly. For example, suppose Bob has a distinct earlobe. This and only this information would be enough to correctly recall who said "The sea looks red today". Another alternative is that we can simply ask, "Was "*The sea looks red today*" said by a man?" But there are two problems with this. First, only a small subset of all processing outputs are going to rise to the level of conscious awareness (Kurzban, 2010, 2011; Kurzban & Aktipis, 2007), which means this method will miss most of the important information-processing occurring within the mind. Second, among the small set of processing outputs that do arise to conscious awareness, all of these are vulnerable to volitional filtering on the part of participants, meaning that participants may not be truthful about what they notice and remember about a particular event for reasons of social-desirability. For example, participants may falsely report not noticing someone's sex, race, or attractiveness, because they think noticing these things in others reflects poorly on themselves. Ideally, then, what we'd like to have is a method that captures the full suite of both conscious and non-conscious information-processing outputs, and a method that is immune to self-presentation biases. This is exactly what using confusions between source-tags accomplishes, and this is exactly what the "Who Said What?" paradigm does.

In the paradigm, multiple people, each saying a different thing, are presented to participants. Participants are not asked to explicitly report what source-tags are attached to each statement. Rather, a situation is set up in which participants are likely to make source tag confusions: participants are asked to try to match which speaker said each of the different statements. This task becomes exceeding difficult with multiple speakers saying multiple statements. The beauty of this method is that the outcome of the memory retrieval process is based on the relative similarity among all of the dimensions processed in the stimuli. For example, if the adaptations in the heads of participants did notice and process Bob's sex, and the sex of the other speakers, then all else equal, the memory retrieval process will produce source-tag confusions between speakers who were similar along that dimension. That is, participants will be more likely to attribute what one man said to another man, and less likely to attribute that statement to a woman.

No matter how minor or major the particular dimension we are interested in is (for example, it may be as obvious as a person's sex, or as minor and apparently imperceptible as a person's earlobe shape), we can then examine how willing participants' memories are to confuse one person with another person. By presenting multiple people, each saying multiple things, and manipulating how many speakers are similar versus different along a particular dimension, we can quantify to what degree that dimension has been part of the information-processing outputs generated. If that dimension is part of the information-processing output, we will see a higher than chance level of within-category confusions (e.g., within sex or earlobe shape), compared to between-category confusions, because it will be tagged in memory. Importantly, because the measure does not depend on explicit reporting of what participants have noticed, and because the memory procedure involves a cumulative pair-wise comparison among all of the processed attributes in the stimuli, rather than just those that rise to level of consciousness, the paradigm is much more sensitive than an explicit measure.

This paradigm has become an important tool for assessing functionalist questions about the human mind. In particular, and because it is so sensitive to information-processing outputs, the paradigm is well-suited to test between competing hypotheses about how cognitive adaptations carve up the social world. Or, to put it colloquially, the paradigm allows researchers to ask of cognitive adaptations: "What is your job…what are you looking for out in world?". For example, the paradigm has been used to test the idea that acquisition-dependent features of language repertoires, such as accents, are picked up by cognitive adaptations which are functionally-independent from those adaptations responsible for tracking alliances and coalitions (Pietraszewski & Schwartz, 2014a, 2014b). It has also been used to demonstrate that adaptations for detecting kin attend to verbal cues of relatedness (Lieberman, Oum, & Kurzban, 2008), and that adaptations for detecting free-riders attend more to the intentions behind particular events, rather than to objective outcomes of the events themselves (Delton, Cosmides, Guemo, Robertson, & Tooby, 2012).

One of the particular strengths of the paradigm is that multiple dimensions of interest can be presented together, but measured independently from one another. For example, participants can be shown people who are members of different races and who also belong to different teams or cooperative groups (e.g., Kurzban, Tooby, & Cosmides, 2001; Pietraszewski, Cosmides, & Tooby, 2014). In these cases, the two dimensions are crossed with one another, which means half of the people shown are on team A, the other half on team B, and each team has half black and half white members. In other words, all four combinations of team and race occur in equal frequencies.

Crossing two dimensions within the paradigm has been particularly useful for testing hypotheses about the functional independence of different dimensions of categorization (or, more precisely, for testing the functional independence of the cognitive adaptations responsible for producing categorization along that dimension). Perhaps the most well-known study of this kind, Kurzban et al. (2001), demonstrated that crossing race with antagonistic team membership reduces categorization by race, whereas categorization by sex remains relatively unaffected when crossed with team membership. This finding, along with an extensive set of follow-up studies (e.g., Pietraszewski et al., 2014; Pietraszewski, Curry, Petersen, Cosmides, & Tooby, 2015), suggests that racial categorization—but not sexual categorization—is a product of adaptations for attending to coalitional alliances. When these coalition-tracking adaptations are provided with information that racial physical features are not reliable alliance cues, the processing of race is inhibited (Pietraszewski, 2016). However, there is a potential problem with these and other studies that feature two crossed dimensions, and this has to do with the correction for differences in baseline probabilities.

1.2. The correction for differences in baseline probabilities

Regardless of whether one or two dimensions are featured in the stimuli presented, all versions of the "Who Said What?" paradigm require a correction for differences in the baseline probabilities of the different kinds of memory errors that can occur. For example, suppose Bob is one of eight different speakers shown in a "Who Said What?" study, in which half of the speakers are male and the other half are female. Participants will be shown an array of all eight speakers during the statement attribution phase, in which they attempt to remember who said which statement. When the statement "The sea looks red today" is presented, participants can respond in one of three different ways. First, they may correctly remember that Bob was the speaker, and