

## EDITORIAL

DOI: <https://doi.org/10.31275/2018/1330>

When I first dipped my toe tentatively into the frigid waters of psi research, back in the late 1970s, one of the big issues of the time was whether the ability to replicate experiments distinguishes—or as philosophers often say, *demarcates*—science from non-science (or pseudoscience). This was a big issue because all too often parapsychological skeptics glibly used that demarcation criterion to bludgeon psi researchers and dismiss them as unscientific. Fortunately, in those days there was some very sensible writing on the subject, particularly from Harry Collins, to whom I was especially indebted when I tackled the topic of replicability myself for the first time.<sup>1</sup> The skeptical position on the issue of repeatability struck me as so lame that I even naïvely expected the debate to be settled rather quickly.

However, because psi researchers often enter the field having little acquaintance with the work that preceded them, and because many critics of that research likewise fail to master the relevant issues, I suppose I shouldn't be surprised that the debate over the nature and importance of replicability still rages. Indeed, little (if any) attention is given to the reasonable points that should have put that issue to rest long ago. Instead, researchers and commentators focus relentlessly—and as usual, inconclusively—on the results of meta-analyses. Some of those meta-analyses are indeed worthy of attention,<sup>2</sup> but (I would say) only in light of the overlooked considerations I discuss below.

So I'd like to review some problems with the still-widely-held view that the ability to replicate experiments is what demarcates science from non-science or pseudoscience.<sup>3</sup> As I see it, that position is both shallow and confused, and the problems with it don't even have the virtue of being subtle. First, the skeptical reliance on the demarcation criterion rests on a naïve conception of the actual importance within science of experimental repeatability. Indeed, experimental repeatability plays little if any role in disciplines (including some physical sciences) whose scientific credentials are not in dispute. Second, it seriously misconstrues how the appeal to replicability works even in those physical sciences where it plays a real role. Third, the received view rests on philosophical confusions regarding the nature of similarity—in particular, the flawed idea that there can be formal, context-independent, criteria for the similarity of two things. And fourth, it rests on confusions over the nature of human abilities generally, and in

particular, the appropriate methodologies for studying them. One could also argue that a fifth problem for this received view is that psi research can in fact point to replicable results. But that last issue must be reserved for another occasion.

### **The Real Role of Replication in Science**

It's clear enough why some people place great emphasis on the replication of experiments, both in parapsychology and in orthodox science. The familiar, underlying idea is that if an experiment E gives a result which replication attempts are unable to reproduce, we have reason to regard E's result as scientifically dubious. And if continued attempts to replicate E fail to duplicate E's result, we have, it would seem, *prima facie* evidence for taking that result to be due to a flaw in E's experimental design, or to experimenter negligence or incompetence, or perhaps even to chicanery. As a rule, then, only experiments whose results can be repeated are considered genuine and reliable. This, clearly, is why some consider experimental repeatability to be a demarcation criterion between science and non-science.

So let's consider first the respects in which the received view's underlying conception of repeatability is naïve. The replicability criterion is obviously borrowed from the physical sciences—but only from some of them (primarily, physics and chemistry). However, experimental repeatability has very little utility in other physical sciences of impeccable credentials—for example, geology and astronomy. Moreover, the received view seriously misconstrues how the appeal to replicability works even in those physical sciences where it plays a major role.

To see this, consider first the abstract question: In what respect(s) can we allow replication attempt E2 to differ from an original experiment E1 and still consider it to be a replication attempt?<sup>4</sup> Obviously, the two experiments can't be alike in *all* respects; they would then be identical, not different experiments. Clearly, E1 and E2 will at least differ with respect to time and/or place of the experiments.

But of course, many other changes will accompany the changes in time and place. These will likely include, for example, differences in the experimental conditions or environment (including inevitable subtle changes in the experimental apparatus required—especially sophisticated, sensitive, and delicate equipment that may continually require fine-tuning), or changes in the actual participants (or just their state of mind). All of these may vary subtly or dramatically from one test to another. But this means that *some* changes between E1 and E2 *must* be tolerated.

But in that case, how is one supposed to determine, *before the results are in*, which differences (if any) matter? What is seldom observed (except

by Harry Collins) is that in *every* science in which experimentation plays a role, it is standard practice to tolerate many differences between original experiments and replication attempts. But that means that scientists in these domains are working with a very loose conception of replication. In fact, scientists who rely on replication attempts don't—and can't—decide, until the results are in, whether the inevitable differences between experiments matter. But that means they can't specify, in advance of conducting a replication attempt, a reliable, much less formal, recipe for replicating the original experiment. Let's look at this in more detail.

Consider first how these observations are true even in the so-called “hard” sciences. In physics, for example, an experiment conducted at laboratory L1 with a certain kind of particle accelerator might be replicated at laboratory L2 with a different design of accelerator. In microbiology, experiments conducted with microorganism M1 in solution S1 might be replicated by studying M1 in a different solution S2 (which may have been more convenient to use, but whose differences are regarded as not making a difference). In fact, even a different microorganism M2 might have been substituted and its difference discounted. And of course, despite the expectations (or at least the hopes) of the replicating scientist, it's always possible that such differences between experiments lead to differences in experimental outcome.

One thing this means is that, *as good science is actually practiced*, the concepts of *similarity of design*, *agreement of results*, or *same result* are both loose and elastic. (We'll return to this point when considering the nature of similarity.) For now, the important point is that the inevitable differences between experiments E1 and E2 will be ignored if their results agree, whereas those same differences might be deemed both relevant and critical if the different experiments yield relevantly different results. But that means *it's not decided in advance—on purely formal grounds—whether the inevitable differences in E1 and E2 matter*. If E2 gets unexpected or undesired results, only then might scientists consider that they were wrong in assuming that the differences didn't matter. But in that case, if E2's results are considered different from those of E1, scientists could easily conclude that E2 wasn't the same experiment as E1, rather than a failed replication. Thus *it can be unclear what the difference is between a failed replication and a different experiment*.

Of course in that case, since differences between E1 and E2 can't be avoided and may lead to a difference in outcome for two experiments, the failure of E2 to achieve the same results as E1 doesn't automatically discredit or even cast serious doubt on E1. Granted, the situation changes somewhat when a series of replication attempts fails to produce the results

of the original experiment. But since all those experiments will likewise differ from each other, it's hardly a straightforward matter to tease out what's responsible for what.

Now in parapsychology, where differences in participants (or their state of mind) may be considerable, and where (thanks to the source of psi problem) we can't be sure who might be influencing experimental outcomes (e.g., the "official" subject, experimenter, onlooker, analyzer), the notions of "same experiment" and "same result" seem especially unclear. But even if we ignore source-of-psi complications, psi research demonstrates the same sort of loose conception of repeatability found in the physical sciences. In parapsychology, E2 may differ from E1 with respect to (for example) the method of stimulating or eliciting a subject's response, providing a subject with feedback, evaluating subjects' responses, the type of interaction permitted between experimenter and subject (including the words spoken and the inflection of those words), and even in the type of response required of the subject.

But this is at most just a difference in degree—not a difference in kind—from what we find in non-behavioral sciences. It certainly doesn't justify the claim that parapsychology is a non-science, or that it's a pseudoscience, or that parapsychology has no repeatable experiments. Methodologically speaking, experimental psi research operates with the same loose conception of repeatability we find in physics, chemistry, and microbiology. And in none of these cases must this reveal a defect in the way the science is practiced. Rather, it's a simple consequence of the inevitable differences between any experiment E1 and any attempted replication E2.

### **Philosophical Problems with the Received View**

To lend a somewhat broader perspective to this discussion, we should also observe that some difficulties in determining when an experiment has been repeated are not peculiar to the scientific enterprise or to the process of experimentation. Rather, they're instances of the more general problem of determining when any sort of event has been repeated. These problems, in other words, concern the general concept of *recurrence*, and even more fundamentally, the concept of *similarity*.

Thus, the question "When does E2 replicate E1?" is at bottom a question about when two experiments count as similar. But the concept of similarity is irreducibly context-dependent. That is, things are not inherently either similar or dissimilar. They must count or be taken as similar or dissimilar relative to some context of inquiry and criteria of relevance. And since no context of inquiry is inherently privileged, that means there's no privileged answer to questions of the form: Is *A* similar to *B*?

Consider for example: Are the movements of an elephant similar to those of a flea? Clearly, there's no privileged correct answer to that question. Sometimes size matters, and sometimes it doesn't. Similarly, does a young beginner's golf swing contain the same movements as the swing of Tiger Woods? Again, there's no privileged correct answer. We might say "yes" if we're comparing golf swings to tennis swings, but not when the focus is on fine differences between the techniques of different golfers. And neither of those perspectives enjoys inherent priority over the other.

Or suppose I try to tell the same joke I heard someone tell the day before. Is the joke I told similar or not to the one I heard earlier? Obviously, it depends on what's relevant to our answering that question, and no criteria of relevance are inherently privileged over the others. Depending on the situation, we might focus on whether my joke made the audience laugh, or whether the words were exactly the same, or delivered at the same speed, or with the same accent or timing, or with the same inflection, or whether my voice had the same timbre as that of my predecessor. The point should be clear: Similarity is not a *static* two-term relation obtaining inherently between the things taken to be similar. Rather, similarity exists *only* with respect to variable and shifting criteria of relevance. It can only be a *dynamic* relation holding between things at a time and within a context of needs and interests.<sup>5</sup>

Likewise, in scientific experimentation, whether E2 replicates E1 is not strictly a function of the formal, much less antecedently specifiable, features of the two experiments. We've already observed that replication attempts will inevitably differ in some respects from the original experiment. But we also noted that in the ordinary course of the established sciences, some of these differences—for example, the mental states of the experimenters, and differences in the equipment used—tend to be discounted when a replication attempt is deemed successful. However, in parapsychology the very same sorts of differences are regarded as potentially relevant to the experimental outcome, although even in parapsychology such differences might also be discounted when replication attempts are deemed successful.

It's also worth noting that there are further complications in deciding what counts as the *same result*. For example, one can raise legitimate questions about what counts as an appropriate level of significance from mean chance expectation. And in parapsychology, a familiar nagging issue is whether extra-chance *negative* results (psi-missing) can be allowed to replicate positive scoring in an earlier experiment. These complications reinforce the points already made about the loose way in which the concept of replication is inevitably used in science, and they needn't be considered in more detail here.<sup>6</sup>

### Human Abilities

Because psi experiments presumably study ostensible abilities of their subjects, it obviously matters what sort of endowment psi might be, and whether the methodologies used to study it are appropriate to those endowments. But then, it's important to note some relevant, but typically unheralded, matters regarding human abilities.<sup>7</sup>

The first point to note is that the notion of a human ability (like the concept of replication) is extremely loose and elastic, covering an enormously wide terrain. In one appropriate and also very common use of the term, "ability" can stand for rudimentary and more or less universal human (or organic) endowments. For example, we can speak of someone's ability to laugh, experience fear, express aggression or compassion, or merely breathe, blink, or move the muscles in one's arm. In this sense of the term, an ability needn't be any kind of proficiency or skill, or disposition to exhibit such a proficiency.

But the term "ability" can also denote various degrees of competence or mastery—for example, when we speak of a person's ability, to learn a new language, carry a tune, hammer a nail, or control pain through self-hypnosis. And of course it can also denote competencies requiring great mastery, as in the ability to play professional-level tennis, write a string quartet, dock a space capsule, read an orchestral score, or solve quadratic equations. "Ability" in this sense seems nearly synonymous with what we usually mean by "skill."

But if we're to have a nuanced, general account of human abilities, we must also consider some other endowments, likewise unevenly distributed among humans, but which we would probably not want to label as skills. Consider, for example, the ability to fire an employee, express sensuality, speak in front of an audience, inspire loyalty in others, remain hopeful in the face of adversity, manipulate others through guilt, and laugh at oneself.

A moment's reflection on the examples above should make it clear that the term "ability," like most ordinary language expressions, has no single and preferred—much less clear and unambiguous—meaning. That's one reason why there's no interesting set of properties shared by all the things we consider abilities, and which distinguish abilities from non-abilities.

At best, the different senses of "ability" merely identify useful points on a continuum of human endowments (ordered roughly in terms of complexity and refinement). And the reason this matters is that it reveals why laboratory research in parapsychology is almost ludicrously premature. It highlights the fact that researchers have no idea what kind of organic function they're trying to investigate. Not only are we ignorant of psi's

finer-grained features, we don't even know what its natural history might be—for example, whether it has an evolutionary role or primary or overall purpose or function (although there is no shortage of speculation on these matters<sup>8</sup>).

Of course, there's no reason to think that psychic phenomena occur only for parapsychologists, much less only when those parapsychologists set out to look for them. After all, a major motivation for conducting formal studies is that we have evidence of psi occurring spontaneously in life. But since we're a very long way from understanding the nature and function of everyday psi, we don't know whether psychic functioning resembles musical or athletic abilities in its variability, or whether it's a brute endowment such as the capacity to see or to move one's limbs. Obviously, then, in the absence of this rudimentary knowledge, we have no idea whether (or to what extent) our experimental procedures are even appropriate to the phenomena.

To see this, compare our knowledge and study of psi with our knowledge and study of memory. Memory is something we *can* study formally to some extent. But we have some idea how to proceed because we're already very familiar with the many and diverse manifestations of memory in daily life. Or compare our knowledge of psi with our knowledge of the ability to be witty. It's because we're familiar with the latter that we know we *can't* adequately study it experimentally. Or again, a tennis player's ability to return serves is something that—unlike everyday psi—we can systematically and easily examine in real-life, relevant settings. In fact, we can study that ability pretty much on demand, and from virtually anyone who claims to have the ability.

It should be obvious, then, that different abilities, as a rule, demand different modes of investigation. We wouldn't examine mechanical aptitude the same way we investigate the ability to produce witty remarks, the ability to baby-sit, the ability to design and install a patio, the ability to learn a new language, the ability to empathize, or the skill of playing football wide-receiver or soccer goal-keeper. Similarly, techniques appropriate to studying those abilities will differ from those suitable for examining rudimentary endowments, such as the capacity to blink, swallow, utter sounds, or dream.

Furthermore, for most human abilities, it's hard to pin down what, exactly, we need to look at. Consider, for example, the ability to compose music. Clearly, that ability can be expressed in many ways. Many composers notate their compositions; others lack that ability. Some composers have absolute pitch, some only relative pitch, and some neither. Some compose directly onto paper, while others need a piano or some other instrument. Some work best with large forms; others don't. Some write especially well or idiomatically only for certain instruments; others don't have that limitation.



Some have a keen ability to set words to music; others lack that sub-ability. Some are especially adept at harmony, rhythm, or instrumental color, and those specialties likewise take different forms and manifest in different degrees and combinations. But then there should be little temptation to think that compositional ability allows many useful generalizations. And there's no reason to think this case is unique; the same is obviously true, for example, in the case of athletic ability, or comedic ability.

What we do know is that people who possess a general ability may exhibit it in various ways and to varying degrees. The differences have to do with the subsidiary abilities or skills they possess and the manner in which they possess them. The moral here should be obvious: At our current level of ignorance, we're in no position to say that psychic functioning is an exception to this rule.

In fact, one can argue plausibly that the manifestation of psi is as deeply idiosyncratic and variable as any other ability. Psi-conducive conditions may be as personal and individual as the conditions people find amusing, or erotic. Most subjects don't do their best under intense pressure or when the stakes are high (say, during a live television demonstration), but a few excel under those conditions and even relish the challenge. And some may be able to demonstrate psi only in the presence of select others—for example, investigators they find especially supportive or agreeable, just as most people can sing or express sensuality only in the presence of those with whom they feel personally safe. Second, the subjective experience of exercising psi varies widely—for example, whether ESP is accompanied by vivid, familiar, or any imagery. And third, the range and specificity of the ability may also vary idiosyncratically. For example, one might be good at psychokinetically influencing small objects but not at affecting random event generators in computerized experiments. Or, one might be good at remote-viewing shapes but not technical details, or colors but not smells, or medium-sized objects but not words on paper (Pat Price, notoriously, was distinctively [if not uniquely] good at this latter task). In fact, this type of ESP variability would parallel a familiar feature of more ordinary perceptual differences. Some are particularly good at (say) discriminating colors but not sounds, detecting subtle differences in wines or chocolates but not in audio components, or noticing eye color but not manipulative behavior.

Moreover, many (and perhaps all) abilities are highly context-dependent and can be expressed or studied properly *only* under quite specific conditions. For example, we can evaluate a tennis player's ability to return serves only under physically and psychologically challenging game conditions. Similarly, a pianist's ability to play the "Waldstein" Sonata, or a comedian's ability to be funny, varies with confidence level,



audience attitude, personal distractions, and so on. So if psychic functioning is analogous to these sorts of organic endowments (as many think and as both experimental evidence and anecdotal reports suggest), then we'd be entitled to say that not everyone is psychic, that some are more psychic than others (enough so to count as "stars" or as gifted), and that not all psychics are psychic in the same way. Needless to say, this can only complicate the process of replicating experiments, both across different subjects, and even with a single subject.

But what if psychic functioning is analogous to elementary capacities? In that case, psi might be as uniformly distributed among humans as pulmonary or reproductive functioning, or as reflexive and involuntary as nursing behavior or fear responses. Moreover, although some lack these familiar capacities or possess them only in attenuated forms, most people have no such limitations. Analogously, the capacity to function psychically might be robust in all but a few individuals. It might also be the sort of thing we do all or much of the time, and the processes involved may be as removed from conscious awareness and control as those involved in digestion or breathing.

However, even if psi functioning is a largely involuntary universal (or nearly universal) endowment, it may still be situation-sensitive to a degree that frustrates attempted replications. After all, our heart rate and digestion, as well as the capacity to sleep, breathe deeply, or ward off infections, can also vary considerably from one occasion to the next. In fact, if the exercise of psi capacities is need-determined (as some have proposed), then it could be analogous to and as variable as the capacity to increase adrenaline flow, or produce endorphins, or the ability to move or respond quickly, act decisively, or be courageous, or cheerful in difficult times, or selfless when a loved one needs to be protected.

Clearly, without some solid grounding, *prior to experimentation*, concerning what sort of human endowment is being investigated, psi researchers can't expect to know, say, whether replicating an experiment with different subjects is even feasible, or whether it's feasible only if the same subject is re-tested, and then only under conditions as similar as possible to those in earlier successful experiments (assuming that can even be determined with any confidence).

So even though parapsychology's replication scorecard may not match that of most physical sciences, and even though that does not undermine parapsychology's status as a legitimate area of scientific inquiry, the field nevertheless remains at an early stage of investigation. Indeed, until we have a more adequate natural history of psi, worrying about replicability may be pointless. In fact, given our current and considerable level of ignorance, one

could argue that research emphasis should still be on proper documentation and vetting of spontaneous or semi-experimental cases.

—STEPHEN E. BRAUDE

### Notes

- <sup>1</sup> See Collins (1976, 1978). Also (and later) Collins (1992). My Collins-inspired discussion was in Braude (1979) and (later) Braude (2002).
- <sup>2</sup> For example, Storm et al. (2017) and Cardeña (2018).
- <sup>3</sup> Some of what follows I covered in an earlier Editorial—in *JSE* 27:1. Evidently, that effort had the usual lack of impact, and so I figure the topic merits another try. Intrepid readers might want to consult that earlier opus for several points not made here.
- <sup>4</sup> I should remind the reader that Collins' original discussions of this issue (and even mine) are considerably more detailed and nuanced. What follows here are simply a few highlights.
- <sup>5</sup> For more on the concept of similarity, see Braude (2014:Chapters 1 and 2).
- <sup>6</sup> But see Collins (1992) and Braude (2002).
- <sup>7</sup> For an extended discussion of this topic, see Braude (2014).
- <sup>8</sup> For some of the best, see Eisenbud (1992).

### References Cited

- Braude, S. E. (1979). *ESP and Psychokinesis: A Philosophical Examination*. Philadelphia: Temple University Press.
- Braude, S. E. (2002). *ESP and Psychokinesis: A Philosophical Examination (Revised Edition)*. Parkland, FL: Brown Walker Press.
- Braude, S. E. (2014). *Crimes of Reason: On Mind, Nature, & the Paranormal*. Lanham, MD: Rowman & Littlefield.
- Cardeña, E. (2018, May 24). The Experimental Evidence for Parapsychological Phenomena: A Review. *American Psychologist*. Advance publication. <http://dx.doi.org/10.1037/amp0000236>
- Collins, H. M. (1976). *Upon the Replication of Scientific Findings: A Discussion Illuminated by the Experiences of Researchers into Parapsychology*. Paper presented at the First International Conference on Social Studies of Science, Cornell University, November.
- Collins, H. M. (1978). *Science and the Rule of Replicability*. Paper presented at the AAAAS Symposium on Replication and Experimenter Influence, Washington, DC, February.
- Collins, H. M. (1992). *Changing Order: Replication and Induction in Scientific Practice*. Chicago: University of Chicago Press.
- Eisenbud, J. (1992). *Parapsychology and the Unconscious*. Berkeley, CA: North Atlantic Books.
- Storm, L., Sherwood, S. J., Roe, C. A., et al. (2017). On the correspondence between dream content and target material under laboratory conditions: A meta-analysis of dream-ESP studies, 1966–2016. *International Journal of Dream Research*, 10(2):120–140.