Editorial

On the Acceptance of New Experimental Findings in Science: Is Replication the Key?

There is an unusual article in this issue of the *Journal*. It is unusual in that the authors explicitly attempted to "replicate" previous results and failed to do so. Yet, here it is—published. The paper was written by Marleen H. M. de Groot and Benjamin Rusak and is entitled "Responses of the Circadian System of Rats to Conditioned and Unconditioned Stimuli." In it the authors describe experiments designed to "replicate" results previously described by Shimon Amir and his coworkers. The results obtained by the latter authors led to the conclusion that neutral stimuli, such as puffs of air, can be made to induce entrainment, or the related phenomenon of phase shifts, by their repeated association with entraining light pulses. Thus, the authors suggested, light can function as an unconditioned stimulus (US) and air puffs can function as the conditioned stimulus (CS) in a classic Pavlovian learning paradigm with entrainment as the response. The authors of the present paper, however, were unable to demonstrate results supporting the conclusion that entrainment can be a learned (conditioned) response. Their paper was extensively reviewed before acceptance for publication; comments were received from five reviewers and it was revised twice.

We felt that the paper itself, the scientific question it addresses, and the broader issues it raises are sufficiently important to warrant highlighting and supplementing the paper in this issue of the *Journal*. To do so, we have added pieces at three levels: a "Technical Comment" from Andreas Arvanitogiannis, Jane Stewart, and Shimon Amir with a "Response" from Benjamin Rusak and Marleen H. M. de Groot; a "Perspectives" provided by Serge Daan, addressing the history and importance of the relationships, such as they are, between learning and entrainment; and this editorial, addressing the vagaries of experimental replication and their role in science.

The importance of replication in science receives a lot of lip service and is commonly considered critical to the scientific process. In fact, we attempt it less often than one might expect and publish the results of such attempts still less often. There are good reasons for this.

First, neither failure nor success in replicating results leads to clear conclusions. Technical details of procedure can easily make attempts at replication fail. This is especially serious in behavioral work, where an old maxim states that "in any well-designed experiment, rats do exactly as they please." Emphasis on exactitude and nuances of procedure may be appropriate, but it suffers the risk of regress into virtually infinite detail and may, simply, be misplaced. Determining the relevant variables for replication is coextensive with determining the properties of the phenomenon in question (as emphasized by Arvanitogiannis et al.) and the validity of the claimed finding. On the other hand, artifactual and "wrong" results may well be reproducible. Indeed, an old maxim from chemistry states that "experiments always come out exactly as they should," though the critical variables may not be what the experimenter intended them to be.

Second, we all want to, and are expected to, make original contributions. Grant applications that aim primarily at replicating the work of others do not get funded. They lack novelty and are deemed second-rate. Generating a body of work that replicates the work of others is not a good way to gain tenure and scientific accolades. Often, we *must* repeat previous work in order to proceed to the next step, but we then aim for and emphasize *extension* of previous work—the new contribution rather than the repetition.

Third, many results aren't worth replicating; confirming or denying their validity would not much affect our confidence in the paradigm they support. Sometimes attempts to replicate the work of others would just take too long, or would be too costly in terms of time and resources, to be worthwhile.

Sometimes we do decide whether a result is valid on the basis of repeated attempts at replication, with variations, and the build-up of consensus. The fate of erroneous, nonreplicable results is to fade from atten-

JOURNAL OF BIOLOGICAL RHYTHMS, Vol. 15 No. 4, August 2000 275-276

tion. Such consensus is often informal, conveyed by word-of-mouth rather than publication, and therefore hard on those dependent on the literature, where the accumulated data are rarely published. Despite the fact that consensus may be reached in this manner, when we don't know the correct outcome, we cannot properly judge validity by replication versus non-replication. Nonetheless, we do manage, collectively, to reach the conclusions and to obtain the scientific benefits often attributed to replication. Replication of data merely indicates reproducibility—an inadequate criterion for validity as mentioned above. It is the *meaning* of the result that counts—its implications and the conclusions to be drawn. It is *confirmation*, or at least support, of that meaning that we seek. "Confirmation" may seem only slightly different from "replication," but it is a critical difference. Resolution of replication issues often comes only from *outside* the original experimental result in question, from some other experimental design that provides an easily reproduced or compelling result, and *its* implications. Such use of different experimental designs is the approach undertaken in the de Groot and Rusak paper.

In our view, the most important criterion for ultimate acceptance of an initially controversial result, the one that provides the confidence in its validity that even reliable replication cannot provide, is its ability to provide a springboard to new science. This is an extension of the emphasis on *meaning* above and goes beyond confirmation of the finding or its immediate implication. It also relates to the question of how much effort is worth expending to confirm a given result. We gain increasing confidence in a result when it turns out we can build on it, when its implications lead to, and bear the weight of, answers to new questions, new findings, new insights, and new perspectives. This could be called the "foundation test" (Can we build on it?) or the "shoulders of giants test" (Can we see farther from it?). These phrases, while conveying the criterion being emphasized, are also a bit misleading. We do stand on the "shoulders of giants" and on the "foundations" built by others, but these images are too static and fixed. More often we stand on, and jump up and down on, and participate in the modification and shoring up of, the *platforms* built by our peers and predecessors. We squabble and scramble over these platforms, and test and replace their planks and beams and railings, wrangling all the while over the design of the scientific structures to be built on them. If they do not hold, we modify them or replace their components or limit their size or dismantle them and rebuild or abandon them.

We think the issue at hand is meaningful and bears important implications. For example, the question arises whether entraining light signals are reinforcers. Do animals and people "prefer" to be entrained? Will they "work" for it? We presume that social factors and time cues can entrain, or affect entrainment of, humans—witness the trouble taken to isolate human subjects from time cues in circadian experiments. Is this really so, and if so, must such entrainment be learned by association with unconditioned entraining stimuli such as light? In another direction, we know that light is aversive to nocturnal rodents. Can it also be a positive reinforcer, insofar as it entrains? If so, the relation between these two effects of light may be complex, and initial studies might be simpler in diurnal species. Other interesting implications and relations between entrainment and learning are addressed, in context, in Serge Daan's "Perspective."

Neither the current proponents of learned entrainment nor the authors who question their conclusion want to be wrong, nor, in this case, does either group want to cast aspersions on the skill or rigor or acumen of the other. Both groups would prefer to sidestep this controversy and both seek common ground on which to reconcile, or at least to resolve, their differences. Although the principals involved find the controversy uncomfortable, they have each called attention to, and will have contributed to our understanding of, important scientific questions.

Martin Zatz
Editor