INTERNATIONAL JOURNAL OF COMPARATIVE PSYCHOLOGY

2018, 31
Charles Abramson
Special Issue Editor
Peer-reviewed

Replication and Pre-Registration in Comparative Psychology

Michael J. Beran

Georgia State University, U.S.A.

There is growing interest and pressure to find ways to address the so-called "replication crisis" in psychology and other areas of science. This includes increasing transparency and implementing good practices in all areas of experimental research, and in particular to promote attempts at replication. Comparative psychology has a long history of efforts to replicate and extend previous research, but it is often difficult to do this when highly specialized methods or uncommon species are being studied. I propose that comparative researchers make greater use of pre-registration as a way to ensure good practices, and I outline some of the ways in which this can be accomplished.

There has been a lot written about the replication crisis in psychology and other areas of science (e.g., Ioannidis, 2005; Kerr, 1998; Lilienfeld, 2017; Maxwell, Lau, & Howard, 2015; Shrout & Rodgers, 2018; Stroebe & Strack, 2014). There is also increased attention about other closely related issues such as increased transparency in the social sciences, and increased efforts to encourage pre-registration of studies (e.g., Hagger & Chatzisarantis, 2016; Nosek & Lakens, 2014) and to find ways to reward replication (e.g., Koole & Lakens, 2012). As comparative psychologists, I think we need to be front-and-center in this discussion, for a number of reasons. I recognize that this is not as easy as it sounds, and thus I am somewhat hesitant to make pronouncements about what a field "should do," although it is reassuring to see that others have made similar pronouncements (e.g., Stevens, 2017). This special issue offers a unique opportunity to present some arguments for what comparative psychology could do, and perhaps should do, and also to discuss issues that we might consider as a field as we navigate the current political and academic climates regarding use of animals in research and regarding replication and transparency in scientific practice.

Is there a "replication crisis" in comparative psychology?

Simply put, we do not know. The more complicated answer is "yes and no." I'll start with the simple answer, which stems from the fact that a substantial number of reports in comparative psychology, just as in other areas of psychological science, are never reproduced. Or, if they are reproduced, they are not published. This can happen because of a number of factors, some of which should be more concerning to us than others. So, we do not know what would or would not replicate in many cases, and we do not know what has already been done, but not published.

I will start with an example from my own career. I have, on at least two separate occasions, clearly documented the reverse-reward contingency effect in the chimpanzees I have worked with at the Language Research Center. The RR task[1] consists of presenting animals with two choice sets, usually of preferred food

[1] Or, if I could, the "Boysen effect," in recognition of Sally Boysen's work in establishing this test paradigm.

Please send correspondence to Michael J Beran, Department of Psychology and Language Research Center, Georgia State University, Atlanta, GA, 30302. (Email: mberan1@gsu.edu) https://doi.org/10.46867/ijcp.2018.31.01.09

differing in the numbers of food items in each set (Boysen & Berntson, 1995). The animals are trained (or often naturally) point toward one of those sets. The trick is that the chosen set is then taken away, or even given to a partner animal, and the *unchosen set* is the one the animal gets. This is a very difficult test for chimpanzees to master, even after quite a bit of experience, at least when real food items are presented. When symbols are introduced, though, they can learn to choose against the better symbol in order to get the better reward outcome (e.g., Boysen, Berntson, Hannan, & Cacioppo, 1996; Boysen, Mukobi, & Berntson, 1999).

The RR test has been given to a lot of other species, as well, with mixed results depending largely on the specific designs used (e.g., capuchin monkeys: Addessi & Rossi, 2010; squirrel monkeys: Anderson, Awazu, & Fujita, 2000; lemurs: Genty, Chung, & Roeder, 2011; great apes: Uher & Call, 2008; Vlamings, Uher, & Call, 2006). The point here is that this is an area in which most of the published papers were not replications of the basic effects so much as efforts to try to get successes from animals. This has been true even in my laboratory, where we eventually came to find some important aspects of what might determine the consistent failures of chimpanzees (Beran, James, Whitham, & Parrish, 2016). Well before conducting that study, I had seen clear failures by my chimpanzees that nicely matched what Boysen and her colleagues had reported. But, I never published, or even submitted, those results, largely because I had no expectation that such efforts to show what someone else already had shown would be valued.

In the case of the RR task, other teams used other species and other approaches such as only rewarding points to smaller amounts, to build a large literature (see Shifferman, 2009, for a review of these approaches and outcomes). But, the point is that this is not always the case in comparative work, especially work on comparative cognition. This leads to the second, more complicated answer. Comparative psychology has, in many areas, been an absolute beacon for replication efforts. For decades, it was a research tradition in comparative work (largely with pigeons and rats, but also some other species), to perform an "Experiment 1" that largely replicated whatever study had inspired the current work, and then to conduct the additional extensions of the replication to add new data to an area. This convention was fairly easy to follow because apparatus was highly consistent across laboratories, and training routines, data collection, and analyses were clearly reported and easy to replicate. And, one can see shining examples of areas in which replication and extension occurred through many years (e.g., intertemporal choice experiments: Ainslie & Herrnstein, 1981; Green, Myerson, Holt, Slevin, & Estle, 2004; Hayden, 2016; Logue, 1988; Marshall & Kirkpatrick, 2016; Marshall, Smith, & Kirkpatrick, 2014; Rachlin, 2000; Tobin & Logue, 1994). However, in many cases comparative researchers often work with unique groups of animals, using customized, one-of-a-kind apparatus, and in many cases test subjects for whom past training experiences are directly related to the design of new experiments. These approaches offer the chance to provide new insights, and new discoveries, but they do not readily lend themselves to easy replication by other laboratories, and so often one has to rely either on that laboratory to self-replicate in future work, or to accept the results of a single experiment (or small number of experiments) as being reflective of "what animals do.[2]" In some cases, replications might be possible, but often cannot be exact, and instead take the form of conceptual replication (see discussion in Agrillo & Miletto Petrazzini, 2012).

What I hope will happen is that those who have the means and opportunity will engage more in replicating the first reports of new phenomena with animals. Perhaps the emerging, broad emphasis on replication in the social sciences will allow us to propose projects that include a replication first, and then the extension to something new. Or, perhaps there can be more efforts by laboratory groups or undergraduate students in a course on animal behavior or learning with a laboratory section to solely focus on a replication

[2] Please note this is not a criticism, but a comment. Most of my own research falls into this category.

effort, so that we might build more confidence in our reported findings. I realize this is not going to occur in all instances; not everyone has access to chimpanzees, or parrots, or elephants, but I think a good step would be to find ways to reward replication efforts when they do occur and ensure that those efforts are present in our literature.

Pre-registration of comparative studies

Another area in which I think comparative psychology could be at the forefront of progress in the social sciences is in pre-registration of reports. The idea behind pre-registration is to write the rationale, experimental design, methods, and planned statistical analysis for a project before collecting any of the data. You then submit that document to a repository so that one "pre-commits" to that plan before data come in, and there might be the temptation to "p-hack" or otherwise adjust the methods, the trial counts, the conditions, or the analyses to "massage" a significant finding from the project (e.g., Ioannidis, 2005; Simonsohn, Nelson, & Simmons, 2014).

Although in principle any area of psychological research can pre-register studies, many forms of comparative research are ideally designed for this. Consider, for example, research that is conducted on quantity comparison. Whether conducted using manual tests with food items (Beran, 2001, 2004), or computerized tasks that carefully control stimulus features (e.g., Brannon & Terrace, 2000), the basic design of these studies is that animals will make choices from arrays, and that those arrays will present specific combinations of quantities. Then, the analysis will examine performance as a function of various properties of those comparisons, such as their relative quantitative differences or the ratio of those sets to each other. These designs easily could be pre-registered for use with new species, or as part of replications of work conducted already with some species. Other examples include maze tasks, classification tasks, concept learning tasks, perceptual discrimination tasks, and many others. In these areas, and others, we know fairly well what designs work best to allow animals to show us their best possible performances.

A concern about pre-registration among those who work with animals is that new approaches would rarely lend themselves to pre-registration. Because we cannot give verbal instructions to animals, they must learn task demands through trial and error, and in many cases this requires a number of changes during pilot testing. So, one might start with a new method and then tweak that method to find an approach that will engage an animal subject and allow it to show its best capacity for the task. This would not allow for pre-registration. I agree that this is a "problem," but I think that one could pilot a new approach with a small subset of subjects, and then propose a pre-registration before collecting additional data with other subjects once the method was appropriate, but before enough data were collected to answer the experimental question. Or, it may be the case that the first efforts in a new area of inquiry are not pre-registered, but this does not mean that subsequent work would not be amenable to such pre-registration.

There is an upside in this effort that is often overlooked. I suspect that many who are reading this are thinking of elegant designs they have created, with intriguing hypotheses about what animals might do (or might not do), and then conducted those studies proficiently, and had good data to analyze. But, the results were the "shortcoming" in that they may have been ambiguous, or gone against the "established knowledge" and thus ended up in the so-called "file drawer" of unpublished experiments. In many cases, this likely happened after extensive use of resources (time, effort, and even animals' lives), none of which ended up being part of our literature, even though those data could have value at some future time. This "file drawer" problem is particularly troubling when you consider that a subsequent team might also expend the same time and

resources (and even animal lives) pursuing a question they think is novel, when in fact it has been investigated already. For these reasons, I find the idea of peer-reviewed, pre-registered reports compelling. I currently serve as co-editor of another journal, *Animal Behavior and Cognition*, where we accept such submissions. Researchers can submit their full plan, with background, design, methods, and analyses, for full peer review. If reviewers agree that the rationale, design, and analyses are appropriate to answer an important question, *Animal Behavior and Cognition* will commit to publishing those results along with the pre-registered paper no matter how the results turn out.[3]

The benefits of peer-reviewed pre-registered reports include allowing for publication of good science that otherwise results in data that may be difficult to interpret or may not fit within established and accepted ideas about animal behavior. When such "odd" or counterintuitive data emerge under the historical system in which many of us trained, they tend not to enter the literature even though the ideas and approaches were excellent. Add to this a scenario in which failed replications may be blocked from publication by defensive reviewers or editors, and you can see the problem. In other cases, good science is rejected for publication because the results were not the "first," "best," or most newsworthy outcomes, thereby hindering the accumulation of publicly available knowledge and creating a situation in which animal time (and perhaps most concerning, animal lives) are "lost" in the sense that the resulting contradictory data never offer any value to anyone outside that laboratory (unless those authors are somehow later asked about failed studies for a meta-analysis or similar effort). From the perspective of ethical use of animals in research, this seems inappropriate as an outcome of the use of animals as well as a problem for the generalizability of our comparative results.

Another example is useful here. For the past few years, we have examined a particular visual illusion called the Solitaire illusion (Agrillo, Parrish, & Beran, 2014; Parrish, Agrillo, Perdue, & Beran, 2016). This is a robust illusion for most humans, in which they see colored dots presented more centrally in Figure 1 as being more numerous than the colored dots presented more peripherally, even though in both arrays there are equal numbers of white and black dots (Frith & Frith, 1972). Despite evidence that nonhuman primates show many of the same visual illusions seen by humans, they do not appear to show the Solitaire illusion, or at least barely seem to see it. We have "kept at" this illusion for some time, working under the assumption that no one else had really looked at this, and even now we continue to work on this under the assumption that perhaps we have not found the right way to present the illusion to best see evidence that nonhuman primates are susceptible to it. If we knew, prior to starting this work, that two, or three, or four other laboratories had, at some point, tried this and found that monkeys and apes had not really shown this illusion, we could have 1) trusted our initial "null" results more, because they would be meaningful failures of animal perception that replicated previous failures or 2) avoided using animal time for this project at all. I am not saying other data exist that show no Solitaire illusion in animals, but I am saying that if such studies had been conducted in the past, it is highly possible that they could not find a home in a publication because they were considered "null results."

[3] Of course, this assumes that the authors conduct the research as outlined in the accepted pre-registration, can collect the proposed data in the way outlined, and perform the analyses that they proposed. Deviations can be tolerated, if they add substantially to the report, are also peer-reviewed, and if they are clearly highlighted as being post hoc to the pre-registration.

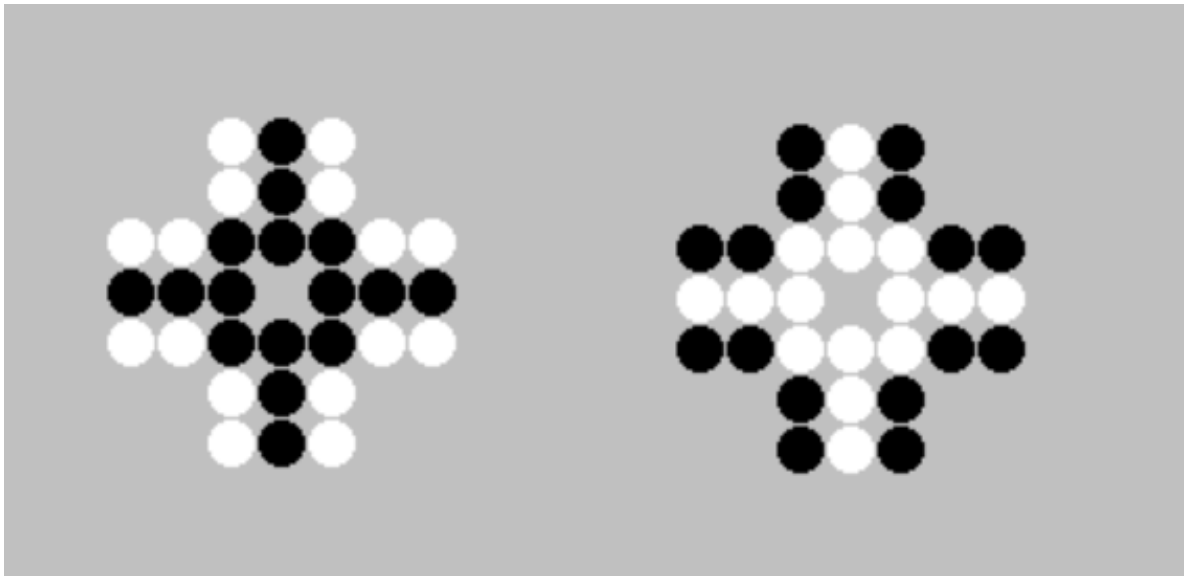

Figure 1. The Solitaire illusion. Both patterns have equal numbers of black and white dots, but in each, most people estimate that there are more dots of the color that is more centrally located (black at left, white at right).

Concerns about pre-registration

In the future, pre-registering studies, and having those pre-registrations peer-reviewed, offers a lot of positive benefits. Of course, there are legitimate concerns. One, which I tend to discount, is that pre-registering an idea requires substantial writing and a necessary time delay during review that would not otherwise have to happen if an idea could be taken more quickly to the data collection phase. I discount this because, if this became common practice, the overall writing time would still be equivalent, it would just be that more of it occurred before data collection rather than after. So, it would be more like the masters and doctoral thesis efforts we all have been through before, where we had to convince others that the results were potentially compelling enough to justify the effort to collect the data, or at least put our ideas down on paper. And, as I noted above, one could also pre-register a study without having it peer reviewed, if one did not want the guarantee of publication after the data were collected according to the plan.

The bigger concern is that such peer reviewed pre-registrations might lead to idea-stealing.[4] I agree this is a concern, but steps could be taken. For example, reviewers could sign a statement agreeing not to discuss the ideas with any colleagues and not to undertake research similar to that in the paper for some period of time, thereby granting the authors priority for their ideas. Editors certainly could adjudicate disputes over this, and do so quite objectively given that they would know what was proposed, and also what a reviewer might then do in his or her own lab. In cases of violation, the journal could have a policy for resolving the issue or even instituting formal charges of academic misconduct against a transgressor. So, I think there are ways to protect researchers while still encouraging them to present their best ideas for peer review, and lessen the file

[4] Again, simply pre-registering is not subject to this, because teams could embargo the pre-registration documents for some period of time, giving them the chance to collect their data.

drawer problem where good data might die lonely deaths in someone's "Unpublished" folder on their computer.

Summary

Comparative psychology often has been at the theoretical or methodological forefront within psychological science. That opportunity may be presenting itself again, although there is a clear wave of interest in these ideas in other areas of psychology. I am not arguing that all studies should be pre-registered, or that all labs need to do more replications of each other's work. This can be a selective process, and a period of adjustment – to allow researchers the time to figure out how best to do this, but also to give us all time to make this new focus something that we advocate for in our departments and universities as being a sign of excellent scientific practice, worthy of resources and worthy of recognition for students and faculty who engage in this effort. And, ideally, to have these ideas move toward broader value-recognition by national associations and funding agencies. The National Science Foundation has recently called for more efforts at replication, and has offered funding for those efforts, in the hope of generating more reliable science. I hope that comparative psychologists will choose to be a part of this new movement.

References

Addessi, E., & Rossi, S. (2010). Tokens improve capuchin performance in the reverse–reward contingency task. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20101602.

Agrillo, C., & Miletto Petrazzini, M. E. (2012). The importance of replication in comparative psychology: The lesson of elephant quantity judgments. *Frontiers in Psychology*, *3*, 181.

Agrillo, C., Parrish, A. E., & Beran, M. J. (2014). Do primates see the Solitaire illusion differently? A comparative assessment of Humans (*Homo sapiens*), chimpanzees (*Pan troglodytes*), rhesus monkeys (*Macaca mulatta*) and capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, *128*, 402-413.

Ainslie, G., & Herrnstein, R. J. (1981). Preference reversal and delayed reinforcement. *Animal Learning & Behavior, 9*, 476-482.

Anderson, J. R., Awazu, S., & Fujita, K. (2000). Can squirrel monkeys (*Saimiri sciureus*) learn self-control: A study using food array selection tests and reverse-reward contingency. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*, 87-97.

Beran, M. J. (2001). Summation and numerousness judgments of sequentially presented sets of items by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 115*, 181-191.

Beran, M. J. (2004). Chimpanzees (*Pan troglodytes*) respond to nonvisible sets after one-by-one addition and removal of items. *Journal of Comparative Psychology*, *118*, 25-36.

Beran, M. J., James, B. T., Whitham, W., & Parrish, A. E. (2016). Chimpanzees can point to smaller amounts of food to accumulate larger amounts but they still fail the reverse-reward contingency task. *Journal of Experimental Psychology: Animal Learning and Cognition*, *42*, 347-358.

Boysen, S. T., & Berntson, G. G. (1995). Responses to quantity: Perceptual versus cognitive mechanisms in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes*, *21*, 82-86.

Boysen, S. T., Berntson, G. G., Hannan, M. B., & Cacioppo, J. T. (1996). Quantity-based interference and symbolic representations in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes*, *22*, 76-86.

Boysen, S. T., Mukobi, K. L., & Berntson, G. G. (1999). Overcoming response bias using symbolic representations of number by chimpanzees (*Pan troglodytes*). *Animal Learning and Behavior*, *27*, 229-235.

Brannon, E. M., & Terrace, H. S. (2000). Representation of the numerosities 1-9 by rhesus macaques (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes*, *26*, 31-49.

Frith, C. D., & Frith, U. (1972). The Solitaire illusion: An illusion of numerosity. *Perception and Psychophysics, 11,* 409–410.
Genty, E., Chung, P. C., & Roeder, J. J. (2011). Testing brown lemurs (*Eulemur fulvus*) on the reverse-reward contingency task without a modified procedure. *Behavioural Processes, 86,* 133-137.
Green, L., Myerson, J., Holt, D. D., Slevin, J. R., & Estle, S. J. (2004). Discounting of delayed food rewards in pigeons and rats: Is there a magnitude effect? *Journal of the Experimental Analysis of Behavior, 81,* 39–50.
Hayden, B. Y. (2016). Time discounting and time preference in animals: A critical review. *Psychonomic Bulletin & Review, 23,* 39-53.
Hagger, M. S., & Chatzisarantis, N. L. D. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science, 11,* 546–73.
Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine, 2,* e124.
Kerr, N. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2,* 196–217.
Koole, S. L., & Lakens, D. (2012). Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science, 7,* 608–614.
Lilienfeld, S. O. (2017). Psychology's replication crisis and the grant culture: Righting the ship. *Perspectives on Psychological Science, 12,* 660-664.
Logue, A. W. (1988). Research on self-control: An integrating framework. *Behavioral and Brain Sciences, 11,* 665-679.
Marshall, A. T., & Kirkpatrick, K. (2016). Mechanisms of impulsive choice: III. The role of reward processes. *Behavioural Processes, 123,* 134-148.
Marshall, A. T., Smith, A. P., & Kirkpatrick, K. (2014). Mechanisms of impulsive choice: I. Individual differences in interval timing and reward processing. *Journal of the Experimental Analysis of Behavior, 102,* 86-101.
Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist, 70,* 487-498.
Nosek, B. A., & Lakens, D. D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology, 45,* 137–141.
Parrish, A. E., Agrillo, C., Perdue, B. M., & Beran, M. J. (2016). The elusive illusion: Do children (*Homo sapiens*) and capuchin monkeys (*Cebus apella*) see the Solitaire illusion? *Journal of Experimental Child Psychology, 142,* 83-95.
Rachlin, H. (2000). *The science of self-control*. Cambridge, MA: Harvard University Press.
Shifferman, E. M. (2009). Its own reward: Lessons to be drawn from the reversed-reward contingency paradigm. *Animal Cognition, 12,* 547-558.
Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology, 69,* 487-510.
Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143,* 534-547
Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science, 9,* 59-71.
Stevens, J. R. (2017). Replicability and reproducibility in comparative psychology. *Frontiers in Psychology, 8,* 862.
Tobin, H., & Logue, A. W. (1994). Self-control across species (*Columba livia*, *Homo sapiens,* and *Rattus norvegicus*). *Journal of Comparative Psychology, 108,* 126-133.
Uher, J., & Call, J. (2008). How the great apes (*Pan troglodytes*, *Pongo pygmaeus*, *Pan paniscus*, *Gorilla gorilla*) perform on the reversed reward contingency task II: Transfer to new quantities, long-term retention, and the impact of quantity ratios. *Journal of Comparative Psychology, 122,* 204-212.
Vlamings, P. H. J. M., Uher, J., & Call, J. (2006). How the great apes (*Pan troglodytes*, *Pongo pygmaeus*, *Pan paniscus*, and *Gorilla gorilla*) perform on a reversed contingency task: The effects of food quantity and food visibility. *Journal of Experimental Psychology: Animal Behavior Processes, 32,* 60-70.

Financial conflict of interest: No stated conflicts.
Conflict of interest: No stated conflicts.

Submitted: *April 14th, 2018*
Accepted: *April 19th, 2018*