

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/132599/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

McIntosh, Robert D. and Chambers, Christopher D. 2020. The three R's of scientific integrity: replicability, reproducibility, and robustness. *Cortex* 129 , A4-A7. 10.1016/j.cortex.2020.04.019

Publishers page: <http://dx.doi.org/10.1016/j.cortex.2020.04.019>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Editorial

The three R's of scientific integrity: Replicability, reproducibility, and robustness

Robert D. McIntosh^{a,*} & Christopher D. Chambers^b

^aHuman Cognitive Neuroscience, Psychology, University of Edinburgh, Edinburgh, UK

^bCardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, UK

***Corresponding author**

Psychology, University of Edinburgh,

7 George Square

Edinburgh, EH8 9JZ

r.d.mcintosh@ed.ac.uk

1. Introduction

1
2 This volume of *Cortex* sees a remarkable double-publication, from a research group at the
3 Centre for the Psychology of Learning and Experimental Psychopathology, KU Leuven. The
4 two articles take complementary approaches to examine the replicability, reproducibility and
5 robustness of a key finding in the cognitive neuroscience of learning, sometimes known as
6 the extinction-reactivation effect. Specifically, they examine the empirical claims, and the
7 original data, of the first report of this effect in humans, published in *Nature* ten years ago
8 (Schiller et al., 2010). Schiller and colleagues claimed that a brief ‘reminder’ (reactivation) of
9 a conditioned fearful association, prior to extinction training, prevented the return of fear
10 when participants were later exposed to the conditioned stimulus. The idea is that reactivating
11 a memory trace returns it to a labile state, opening a reconsolidation window during which it
12 is possible to re-write the memory with new (non-fearful) associations. The potential
13 application of such a simple, non-pharmacological intervention to treat emotional memory
14 disorders is obvious, and Schiller and colleagues’ findings have attracted extensive academic,
15 clinical and media attention. But that attention has never been directed so intensely at the
16 procedural and empirical detail of the original *Nature* paper, as it is in this volume of *Cortex*.

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

These re-examinations of the reactivation-extinction effect make important contributions to the field, and we are pleased to publish them. In what follows, we highlight the editorial processes and article types that have facilitated these contributions. *Registered Reports* is an article type launched at *Cortex* in 2013, and adopted by more than 240 journals since, in which the rationale and study plan undergo peer review prior to data collection (Chambers, 2013; Chambers, Dienes, McIntosh, Rotshtein, & Willmes, 2015). *Cortex* has published 23 Registered Reports to date, but the article by Chalkia, Zenses, Schroyens, Van Oudenhove and Beckers (2020) is notable for the sheer scale of empirical effort, and the number of twists-and-turns along the road. The second article, by Chalkia, Van Oudenhove and Beckers (2020), is not based on new data, but a painstaking reanalysis of the original data from Schiller et al. (2010). It is the first example of a new article type launching in this volume of *Cortex: Verification Reports* (Chambers, 2020).

2. Persistent attenuation of fear memories in humans: A registered replication of the reactivation-extinction effect (Chalkia, Zenses, Schroyens, Van Oudenhove and Beckers, 2020)

1 Chalkia and colleagues' *Registered Report* has had a fraught journey since its original
2 submission as a Stage 1 plan in November 2016. The initial review process was deceptively
3 smooth, aided by the fact that a direct replication was proposed, so that design choices were
4 determined largely by the original experiment (Experiment 1, Schiller et al., 2010). The
5 major difference, aside from the dropping of a non-essential comparison condition, was that
6 the hypothesis test had to meet the high power requirements of *Cortex Registered Reports*
7 (power $\geq 90\%$), implying a sample size ($n=124$, split between two groups) nearly three times
8 larger than that of the original study ($n=42$, across the same groups). This was no small
9 commitment, given that the study required each participant to be tested on three days, and
10 involved fear-conditioning with mild electric shocks. Stage 1 In Principle Acceptance (IPA)
11 was awarded in February 2017, meaning that the journal committed to publish the final
12 report, regardless of the outcome, provided that the authors followed the agreed protocol.
13
14
15
16
17
18
19
20
21

22 Within one month of starting data collection, the researchers realised that around
23 three-quarters of the first 35 participants tested would fall foul of the preregistered exclusion
24 criteria, as compared with an exclusion rate of just 8% reported in the original *Nature* paper
25 (Schiller et al., 2010). In seeking to understand this discrepancy, they examined a more
26 detailed report of the same procedures, in the *Journal of Visualised Experiments (JOVE)*
27 (Schiller, Raio, & Phelps, 2012), and noted that a slightly different set of exclusion criteria
28 were given there. Invited to clarify this ambiguity, the lead author of these reports stated that
29 the *JOVE* criteria, rather than those in the *Nature* paper, were accurate (Schiller, personal
30 communication, 5 April 2017). This was odd, but it did not offer a solution for the replication
31 attempt, because the *JOVE* criteria would have produced a similarly severe exclusion rate. At
32 this point, the replicating authors considered terminating the study for which they had IPA,
33 and then re-registering with a more realistic set of criteria. Before breaking IPA, however, we
34 advised them to try to resolve a more basic question: how could the exclusion rates differ so
35 dramatically if their study was a direct replication of the original?
36
37
38
39
40
41
42
43
44
45
46
47
48

49 To address this issue, Chalkia and colleagues suspended testing in March 2017, and
50 made a data request to Schiller and Phelps (the lead and senior authors of the original study).¹
51 The data were finally obtained, five months later, in August 2017. The following month,
52 Schiller contacted the journal to advise us that the data had also been posted on the Open
53
54
55
56
57

58 ¹ . *Nature* does not require open data with publication, but it does require that materials, data, code, and
59 associated protocols be provided promptly on request: [https://www.nature.com/nature-research/editorial-](https://www.nature.com/nature-research/editorial-policies/reporting-standards)
60 [policies/reporting-standards](https://www.nature.com/nature-research/editorial-policies/reporting-standards).
61
62
63
64
65

1 Science Framework², and to add some further context. This included an admission that the
2 exclusions from Experiment 1 reported in the *Nature* paper (n=6), related only to participants
3 who had completed all three days of testing, and that a much larger number of unreported
4 participants (around 50% of the total enrolment) had been excluded based on “a judgment
5 call after day 1 or 2 data became available” (Schiller, personal communication, 13 Sep 2017).
6
7 Schiller also informed us that the original research team intended to submit an erratum to
8
9 *Nature* to correct the misreporting (this is now published, though as an “addendum” not an
10 erratum: Schiller et al., 2018).
11
12
13
14

15 At this stage, the priority for *Registered Reports* was to establish, if possible, some
16 objectively-defined exclusion criteria that would allow the replication study to mirror those
17 actually implemented in the original experiment. In communication with the journal, and with
18 the replicating authors, Schiller and Phelps attempted to provide such criteria, but were
19 unable to specify rules that fully captured the pattern of exclusions and inclusions. Instead, a
20 rough approximation, covering 93% of cases, was proposed. This rather tortuous list of 11
21 conditional statements was accepted by the editorial team, and endorsed by reviewers, as an
22 appropriate substitution for the preregistered exclusion criteria. This was considered a
23 necessary correction for misreporting in the original study, to preserve the agreed aim of
24 direct replication. We therefore judged that it incurred no risk of bias and did not invalidate
25 the original IPA. In September 2017, Chalkia and colleagues were able to resume data
26 collection with the amended protocol, retaining their 35 existing datasets, albeit half-a-year
27 behind schedule and with a mountain to climb to reach a sample size of 124. Given the
28 stringent exclusion criteria, they eventually tested 246 participants, completing this Herculean
29 labour in February 2020.³
30
31
32
33
34
35
36
37
38
39
40
41
42

43 Detailed accounts of the anomalies in the definition and reporting of exclusions are
44 provided by Chalkia et al (2020) and Chalkia, Van Oudenhove and Beckers (2020). More
45 relevant to this editorial are some of the key factors that have helped to bring these anomalies
46 to light. Obviously, this owes much to the rigour and tenacity of the replicating authors, and
47 to the integrity of the original authors in acknowledging the shortcomings of their previous
48 practices and sharing their data. But the *Registered Reports* policies have also played a key
49 role. Direct replications are essential to science, yet they are notoriously hard to publish by
50
51
52
53
54
55
56
57

58 ² <https://osf.io/jhu5c/>

59 ³ See Figure S1 of Chalkia et al (2020) for a graphical timeline of these events.
60
61
62
63
64
65

1 traditional routes, which put a premium on novelty. Informative replications are onerous,
2 because they must be sufficiently well-powered that null results can be meaningfully
3 interpreted, which often requires a sample two-to-three times larger than that of the original
4 study (Simonsohn, 2015). The *Registered Reports* format uniquely incentivises such major
5 investments, because the IPA, awarded before data collection, virtually guarantees the
6 eventual publication of results. As it happens, direct replications are very well-suited to the
7 format, because there is (or should be) no ambiguity over the methods to be used, or the
8 hypotheses at stake, and there is always at least one prior study to inform power calculations.
9

10
11
12
13
14
15 *Registered Reports*' insistence on the specification of every procedural detail then
16 puts the methods under more exacting scrutiny than is usually achieved in any other way.
17 Notably, previous peer review, conceptual replications, and meta-analyses of the extinction-
18 reactivation effect, had not uncovered the problems with the original study that have been
19 brought to light by this *Registered Report*. Even the inconsistency between the exclusion
20 criteria reported in *Nature* (Schiller et al., 2010) and those reported in *JOVE* (Schiller et al.,
21 2012) was not noticed until placed under this microscope. We also believe that the formal
22 commitment contained in the journal's IPA provided both impetus and support for the
23 replicating authors to pursue clarifications from the original authors with the persistence
24 required. The independent interest of the journal may have helped to convey that the
25 requirement for clarity was not personally directed, but was merely essential for the
26 *Registered Reports* process.
27
28
29
30
31
32
33
34
35
36
37
38
39

40 **3. Preventing the return of fear in humans using reconsolidation update mechanisms: A** 41 **verification report of Schiller et al. (2010) (Chalkia, Van Oudenhove and Beckers, 2020)**

42
43
44 As a consequence of this process, (most of) the original data are now in the public domain.²
45
46 ^{above} Beyond testing the replicability of the original claim, this created an additional
47 opportunity to examine its computational reproducibility, and to investigate the influence that
48 exclusions based on “judgement calls” may have had. This opportunity also coincided
49 fortuitously with plans for another article type at *Cortex: Verification Reports*. The purpose of
50 this format is to assess the credibility of research conclusions, by testing the reproducibility
51 of original analyses and, where appropriate, their robustness to novel analyses (see the
52 accompanying editorial: Chambers, 2020). Given concerns around reporting standards, and
53 confusion over exclusion criteria, the original data of Schiller *et al.* were an obvious fit for
54
55
56
57
58
59
60
61
62
63
64
65

1 the format. Chalkia, Van Oudenhove and Beckers' re-analyses of these data are published as
2 our inaugural *Verification Report*.
3

4 The details of these analyses, the issues they address, and the further problems they
5 uncover, are fully described in the report itself, and they do not make light reading. The
6 authors repeat the critical analyses from Experiment 1, using the exclusion criteria stated in
7 the original *Nature* paper (Schiller et al., 2010), or the criteria stated in the recent addendum
8 (Schiller et al., 2018), or no exclusion criteria, or the idiosyncratic set of exclusions based on
9 qualitative judgements that the original study actually used. Only the last scenario yielded a
10 pattern of results at all consistent with the conclusions of that paper; and even here the critical
11 interaction to test for differences in the reinstatement of fear between groups was not
12 statistically significant. In recognition of the fact that the study's claims were not based
13 exclusively on Experiment 1, the *Verification Report* also included a re-analysis of
14 Experiment 2 from the original paper. The role of exclusions could not be investigated for
15 this second experiment because data were not available for all excluded participants. But
16 from the data that were obtained, the authors were unable to reproduce the statistical result on
17 which the conclusions depended. A host of other inconsistencies were also identified.
18
19
20
21
22
23
24
25
26
27
28
29

30 The *Verification Report* paints an undeniably bleak picture of this influential *Nature*
31 paper and concludes that the evidence for its claims is unreliable. The highly-powered
32 replication in the yoked *Registered Report* did not find any differential reduction in the return
33 of fear after reactivation treatment. None of this reflects upon the quality of other studies of
34 the phenomenon, which have had mixed outcomes (see Kredlow, Unger, & Otto, 2016). But
35 it does suggest that Schiller and colleagues' (2010) study should no longer be considered in
36 the balance of evidence for the extinction-reactivation effect in humans.
37
38
39
40
41
42
43
44
45

46 **4. Conclusions**

47
48

49 Methodological misreporting and the potential unreliability of key findings are causes for
50 concern in any field. The present example also highlights wider issues around the influence
51 of publication practices on scientific progress. Science is often said to be self-correcting, as
52 indeed it may be when viewed impersonally, and on grand timescales. But at the human scale
53 of research careers, the correction process is neither automatic nor immediate. The higher the
54 profile of the finding, and the more important the topic, the more time and resources may be
55 wasted in following it up, and the more effort may be required to correct it. The epic double-
56
57
58
59
60
61
62
63
64
65

1 publication in this volume of *Cortex* is a visible example of this effort, not unlike previous
2 examples in recent volumes (e.g. Potter, Huszar, & Huber, 2018; see accompanying
3 comments by Arguello, 2019; Chambers, 2019; Hobson, 2019; Huber, Potter, & Huszar,
4 2019; Inzlicht, 2019; Maizey & Tzavella, 2019; Schwarzkopf, 2019; Wall, 2019). Through
5 the process of peer review, we also learned of conceptual replications of Schiller and
6 colleagues' (2010) protocol that were abandoned due to very high exclusion rates. This
7 cannot be a worthwhile use of resources, regardless of whether or not reactivation treatment
8 will ultimately translate into effective therapies for emotional memory disorders.

9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

As scientists, we all make mistakes, and we may trust (impersonally) to science to correct them in the long run. But it is even better if we can head off the most avoidable errors at source, while also embedding mechanisms for checking the reliability of findings that do enter the literature. As gatekeepers of publication and reputation, journals have a heavy responsibility to ensure that the work they publish is as trustworthy and transparent as it can be. This burden certainly falls on the higher-impact journals in a field, where the influence and rewards of publication tend to be greatest. In this spirit, *Cortex* has established three article types, aimed at increasing trust and transparency at each stage of the research process. *Exploratory Reports* exists to encourage the presentation of hypothesis-generating research as openly exploratory, rather than confirmatory (McIntosh, 2017). *Registered Reports* is ideally-suited for hypothesis-testing and direct replication; its editorial processes have been designed to select and publish high quality experiments, regardless of results, neutralising many pernicious sources of bias (Chambers, 2013; Chambers et al., 2015). Now *Verification Reports* is added, as a check on the reproducibility and robustness of prior findings.

Cortex also subscribes to the Transparency and Openness Promotion guidelines (Chambers, 2018). This means that authors of empirical articles are required to share their materials, data and code to the maximum extent permitted by ethical and legal constraints. There are cases where legitimate barriers exist, but digital storage means that open data should now be the norm and not the exception. Solid science is best-served when materials and methods are sufficiently well-specified to allow direct replication, and where the data are freely available for independent checks of reproducibility and robustness. By normalising open practices, we may hope to minimise the less welcome but occasionally necessary fourth R of scientific integrity: retraction.

References

- 1
2 Arguello, P. A. (2019). Market forces influence editorial decisions. *Cortex*, *113*, 363–364.
- 3
4 Chambers, C. D. (2013). Registered Reports: A new publishing initiative at Cortex. *Cortex*,
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- Chambers, C. D. (2018). Introducing the transparency and openness promotion (TOP) guidelines and badges for open practices at Cortex. *Cortex*, *106*, 316–318.
- Chambers, C. D. (2019). The battle for reproducibility over storytelling. *Cortex*, *113*, A1–A2.
- Chambers, C. D., Dienes, Z., McIntosh, R. D., Rotshtein, P., & Willmes, K. (2015). Registered Reports: Realigning incentives in scientific publishing. *Cortex*, *66*, A1–A2.
- Hobson, H. (2019). Must replication attempts be battlegrounds? *Cortex*, *113*, 355–356.
- Huber, D. E., Potter, K. W., & Huszar, L. D. (2019). Less “story” and more “reliability” in cognitive neuroscience. *Cortex*, *113*, 347–349.
- Inzlicht, M. (2019). Transcending humanness or: Doing the right thing for science. *Cortex*, *113*, 360–362.
- Kredlow, A. M., Unger, L. D., & Otto, M. W. (2016). Harnessing reconsolidation to weaken fear and appetitive memories: A meta-analysis of post-retrieval extinction effects. *Psychological Bulletin*, *142*(3), 314–336.
- Maizey, L., & Tzavella, L. (2019). Barriers and solutions for early career researchers in tackling the reproducibility crisis in cognitive neuroscience. *Cortex*, *113*, 357–359.
- McIntosh, R. D. (2017). Exploratory reports: A new article type for Cortex. *Cortex*, *96*, A1–A4.
- Potter, K. W., Huszar, L. D., & Huber, D. E. (2018). Does inhibition cause forgetting after selective retrieval? A reanalysis and failure to replicate. *Cortex*, *104*, 26–45.
- Schiller, D., Monfils, M. H., Raio, C. M., Johnson, D. C., Ledoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, *463*(7277), 49–53.
- Schiller, D., Monfils, M. H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2018). Addendum: Preventing the return of fear in humans using reconsolidation update mechanisms (Nature (2010) 463 7277 (49-53)). *Nature*, *562*(7727), E21.
- Schiller, D., Raio, C. M., & Phelps, E. A. (2012). Extinction training during the reconsolidation window prevents recovery of fear. *Journal of Visualized Experiments*, (66).
- Schwarzkopf, D. S. (2019, April 1). The red thread in the maze. *Cortex*. Masson SpA.

1 Simonsohn, U. (2015). Small Telescopes: Detectability and the Evaluation of Replication
2 Results. *Psychological Science*, 26(5), 559–569.

3 Wall, M. B. (2019). Reliability starts with the experimental tools employed. *Cortex*, 113,
4 352–354.
5
6
7
8
9

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65