**Malte Elson & Ian Hussey**
Psychology of Digitalisation
University of Bern
malte.elson, ian.hussey@unibe.ch

*EXAMINING THE REPRODUCIBILITY AND ROBUSTNESS OF RESEARCH*

**OBJECTIVES |** ERROR is a comprehensive program to systematically detect, report, and prevent errors in scientific publications modelled after bug bounty programs in the technology industry with five major goals: 1) Exploring and testing practical challenges in its implementation. 2) Estimating a benefit-cost ratio of funding error detection. 3) Obtaining robust empirical estimates of types of errors and their prevalence. 4) Surveying, documenting, and increasing accessibility to the digital tools used to detect errors. 5) Fostering a culture that is open to the possibility of error in science, and that embraces a new discourse norm of constructive criticism.

**RATIONALE |** The scientific enterprise, as a human profession not immune to error, has developed several failsafe mechanisms (e.g., peer review) that also reflect one of its basic tenets: Science is self-correcting. However, these mechanisms are not purposely designed to catch errors, and often only do so long after errors have already proliferated in the literature. Just as most errors are coincidental, science currently relies on their coincidental discovery by highly motivated individuals that know how to spot them. Further, error detection as a scientific activity is relatively unappealing as there is little to gain and much to lose for both the researchers whose work is being scrutinised (making cooperation unlikely), and those who do error detection.

**METHOD |** In ERROR, analogous to bug bounty programs in software development, investigators examine published works in journals of psychology and preprint servers (including study materials, data, and code) for errors and receive compensation in form of a base rate and variable compensation depending on the severity of their findings. Similarly, authors who agree to having their work examined this way, and commit themselves to appropriate consequences where necessary, receive compensation if their work proves to be reliable. ERROR selects papers eligible for review based on their importance or replication value, identifies suitable investigators to check the work, acts as a liaison between investigators and authors (and editors, if necessary), acts as an arbiter of the severity of the findings that the investigators report, and collates investigator reports to document the usefulness of error detection techniques and facilitate meta-scientific research.

**COMMUNITY |** The development and rollout of ERROR is accompanied by surveys of research communities, which cover (1) practical and logistical aspects (e.g., expected turnaround time, fairness of compensation), (2) procedure and technology acceptance (e.g., fears of reputational costs, power differences between authors and investigators), and (3) viable alternatives to ERROR's proposed mechanisms against error proliferation. As ERROR's rollout will start in psychology, and later be extended to other social sciences, such surveys can also highlight important cross-disciplinary similarities and differences to be considered.

**SUSTAINABILITY |** A cost-benefit analysis of implementing ERROR on a larger scale will need to consider (1) running costs, (2) counterfactual costs (e.g., papers existing in the literature without having been purposely reviewed for errors), (3) consequential costs (e.g., further grants awarded building on flawed research), and (4) opportunity costs (other ideas not pursued instead of flawed research). Involving important stakeholders is necessary for a sustainable adoption of ERROR: (1) funders, who share responsibility to take measures against wasteful spending, (2) publishers, who share responsibility to implement detection and prevention measures before errors proliferate, and (3) academic societies and associations, who share responsibility to integrate error detection and prevention in community building as part of normal science. The overarching sustainability strategy that will benefit any implementation effort is the automation and democratization of error detection.

WISSEN
SCHAFFT
WERT**.**