Nudging Research Integrity: A White Paper

This white paper is addressed to teachers, group leaders, and heads of institutions at universities. We highlight ways of promoting research integrity by the use of soft interventions or nudges. The use of behavioural insights for promoting research integrity is still in its infancy. Studies in this field are few in numbers (Anderson & Adam, 2014), and the general domain of nudging suffers from uncertainty due to general replicability issues within psychology. Despite these limitations, a realistic appraisal of the scientific results from this literature allows us to highlight a few simple principles that have proved effective and that decision-makers should keep in mind when promoting research integrity: they should strive for interventions that are Easy, Attractive, Social and Timely (the EAST framework), and that increase observability, communicate expectations, and eliminate excuses (Yoeli's three principles).

In what follows, we explain those general principles, and illustrate how they might be applied to promote research integrity. Each decision-maker willing to promote integrity will be able to tailor these principles to specific needs and contexts.

Aurélien Allard & Christine Clavien

Geneva University

1. Introduction

Teaching research integrity is crucial in the 21st century. More students than ever are going to university, highlighting the potential for higher education to serve as a general environment for promoting virtuous citizens. At the same time, recent years have seen an increase in scientific scandals, linked with fraud (Stricker & Günther, 2019) or the lack of replicability of scientific experiments (Errington et al., 2021; Open Science Collaboration, 2015). For instance, the proportion of scientific papers in psychology journals retracted for misconduct has steadily increased from 2000 to 2010, going from a baseline of approximately 0.5 per 10,000 published articles to 2.5 (Stricker & Günther, 2019). While this increase likely indicates the greater visibility of misconduct, it also illustrates the need for improved moral education.

Research integrity is a broad category, which includes several subdomains. Philosopher Heather Douglas offers the following important subdivisions (Douglas, 2014): integrity in promoting cooperation among scientists, integrity in protecting research subjects, and integrity in promoting knowledge. The first category includes norms related to plagiarism and the proper attribution of authorship. The second category includes ethical norms related to the rights of research participants. The third category includes all norms related to the proper uses of scientific methods, and the prevention of fraud and questionable research practises.

Promoting cooperation in science and protecting research subjects have long been core targets of research integrity interventions. Promoting the proper use of scientific methods has recently come to the fore as well, as recent scientific history has highlighted the need for increased openness and transparency in science in the wake of the replicability crisis. The fact that many experiments have failed to replicate both in social sciences (Camerer et al., 2016; Open Science Collaboration, 2015) and natural sciences (Errington et al., 2021) highlights the need for new measures such as the recourse to open data, open materials, and pre-registration, and improved methodological education (Christensen & Miguel, 2018; Nosek et al., 2022). In this report, we highlight interventions touching on all three dimensions of scientific integrity.

This paper highlights the importance of nudges, within the general framework of behavioural insights. Behavioural insights is a catch-all term that includes all interventions shaped by evidence-based understanding of human psychology, including biases and heuristics people use. Within the general context of behavioural insights, nudges are the interventions that have attracted the strongest attention. Nudges are small interventions that do not constrain behaviour or that do not strongly shift incentives faced by participants, but that modify the decision framework so that participants have a greater tendency to choose desirable options (Thaler & Sunstein, 2021). In other words, nudges are soft interventions that guide without coercing. It can be opposed to strict incentives and regulations, which limit the option available to people, and educational interventions, which do not coerce but typically provide information and involve a stronger time or cognitive investment on the part of the participant.

2. Nudges To Promote Pro-social Behaviour

In this white paper, we will present a selection of nudges to promote research integrity. Research in this area is still in its infancy, and we mainly base our propositions on results from research aiming at promoting general pro-social behaviour. Pro-social behaviour refers to any kind of behaviour that benefits other people or society. Examples of prosocial behaviour include paying taxes, performing pro-environmental actions, or giving blood or organs. Research integrity is a specific kind of prosocial behaviour: it includes behaviours that benefit research participants, other researchers, or society, in the context of the production and dissemination of scientific knowledge.

Nudges and behavioural insights can take a large number of different shapes (Sunstein, 2014). However, an attractive summary has been proposed by the British behavioural insight team (Halpern, 2016): we should strive to promote behaviours by designing interventions that are Easy, Attractive, Social and Timely (EAST). Making behaviours easy refers to the removal of bureaucratic obstacles and to the facilitation of the adoption of desirable actions. The importance of attractiveness is self-explanatory. Making a behaviour social refers to the importance of promoting a social context of norms and accountability. At last, making an intervention timely refers to the idea that it should be targeted at the precise moment when participants need it.

Three Nudges Inspired by the EAST Framework

While all four principles are intuitive enough, they can generate nudges that have deep impacts on people's behaviours. Here are three examples of successful nudges based on the EAST framework: promoting adequate defaults, providing reminders, and giving feedback.

The default option	The reminder	The social feedback
A default option refers to the option that will be automatically applied to someone if they choose not to opt out from a specific choice. The idea of the default nudge is that choosing a prosocial default option makes it *easier* to stick with it, and signals that other people care about the pro-social behaviour, thus activating a *social* aspect. For instance, one recent article studied the impact of having a vegetarian or a non-vegetarian meal as the default option at scientific conferences (Hansen et al., 2021). While hardly anyone chose the vegetarian meal when meat was the default, selecting the vegetarian option as the default led to a shift from close to 0% to almost 95% of participants eating vegetarian meals. The strength of this impact is all the more surprising since participants only had to tick a box if they wanted to have a meaty meal instead of the vegetarian option. This shows that nudges can have a tremendous impact in terms of important moral behaviours.	Another nudge that makes adequate behaviours both *easier* and more *timely* relies on sending reminders to participants. For instance, simply sending texts to participants reminding them that they have an upcoming court hearing leads to a halving of accused people not showing up to their court hearing (Fishbane et al., 2020).	A third nudge that has been successful in a variety of cases consists in providing feedback to participants regarding their own performance or practice. For instance, one nudge provided feedback to participants regarding their energy consumption, indicating whether they consumed more or less energy than the average person in their area (Allcott & Rogers, 2014). On average, people who were above the average in terms of energy consumption reacted by reducing their energy use. This nudge works by highlighting *social* comparisons, with people not wanting to degrade the environment more than their peers.

The EAST framework is a general tool to increase any kind of behaviour, including behaviours that are ultimately useful for the participant themselves. However, other frameworks have been developed to target pro-social behaviours specifically. One of the most interesting frameworks concerns the three different maxims promoted by behavioural economist Erez Yoeli: interventions for promoting pro-social behaviours work best when they (1) increase observability, (2) eliminate excuses, (3) communicate expectations (Yoeli, 2018).

The importance of increasing observability stems from the centrality of social factors: participants want to have a good reputation among their peers, so making their behaviour public will cause them to behave pro-socially. Eliminating excuses and communicating expectations both stem from the tendency people have to adopt motivated reasoning in cases of unclear rules: if there is a possibility of justifying one's actions while still choosing behaviours with personal benefits, then people will tend to go for it. Heads of institution and group leaders should thus take the time to clearly communicate expectations regarding the Responsible Conduct of Research, in order to promote the observability of responsible conduct, and to remove the possibility of cheap excuses for misconduct.

Why We Should Remain Cautious Regarding the Impact of Nudges

Nudges have been widely popular in decisional circles ever since the publication of Cass Sunstein and Richard Thaler's book Nudges in 2007. Behavioural insights team have been created in many countries such as the UK, the US, and Australia, in order to improve public policy through the uses of behavioural sciences. However, we should remain cautious regarding the overall impact of nudges, and we should avoid seeing them as a magic bullet. While some nudges undoubtedly work, their overall impact tend to be small. In 2022, economists DellaVigna and Linos made a general review of the impact of nudges in two American Nudge Units, and in the published literature in behavioural economics, as estimated by recent meta-analyses (DellaVigna & Linos, 2022). They found that, while papers in the academic literature estimated that nudges increase a wide range of desirable behaviours by an average of 9 percentage points, the impact estimated by the nudge unit was an increase of 1 percentage point. The gap between the two estimates was probably due to publication bias in the academic literature, as the estimate from the nudge unit incorporated all conducted trials and was unbiased. The average impact of nudges is thus probably very small.

The possible impact of biases in the estimates of the effect of nudges can be seen in the uncertainty regarding one particular type of nudge, the social norms nudges. These nudges assumed that participants would be influenced by knowing that the majority of participants conformed to positive actions (e.g. giving participants the information that "the vast majority of citizens correctly pay their taxes"). While numerous studies showed a positive impact of these interventions, recent replications have generally shown a null effect (Bohner & Schlüter, 2014; Dimant et al., 2020; Silva & John, 2017), revealing the possible distortion in the academic literature. As results accumulate, it becomes increasingly clear that many nudging interventions don't work and those who work may be efficient in some contexts but not in others. As a consequence, researchers have recently started to investigate models that could help predict whether a planned nudge is likely to be effective or not (Callaway et al., 2022).

3. Nudges to Promote Integrity

Based on the principles highlighted above, we provide a series of examples of nudges and interventions that have been robustly shown to have a positive impact on pro-social behaviours. However, we do not claim that any application of these principles in a teaching or research context would have a positive impact. Our suggestions should be seen as a toolbox that practitioners will have to critically evaluate and adapt to their local context.

Practical Case 1: Formal Academic Settings

We can use behavioural insights to promote research integrity within formal teaching contexts, such as classes and seminars on research integrity. The EAST framework articulates principles that have long been familiar to teachers. Making resources easy and attractive, for instance, are two ideas that are probably in the mind of any educator. However, it is important to illustrate how these principles can be applied to the specific context of research integrity.

A general trend for making research integrity training more attractive involves the use of gamification. If students learn about research integrity in the context of a game that requires them to make active choices, they are likely to be more involved in the learning process and to find the educational content more relevant for them. A recent example of a game developed for teaching research integrity in the classroom is the Integrity Games website (*Integrity Games*, 2022). On this website, students can play the role of young researchers having to make choices involving research integrity issues, such as data management problems or plagiarism cases. The website is organized in the way of a game-book, with different choices in the dilemmas leading to different outcomes. The website is intended to be used to stimulate discussions in the classroom.

Beyond games, behavioural insights can also be used to increase the impact of traditional integrity courses. Let's take the timely idea of the EAST framework. Current research integrity training often takes place at the beginning of one's research career, which means that students have to integrate a vast range of "good" practices long before they have conducted any academic research themselves. It might consequently be hard for students to really engage with the material since they have not experienced these issues yet. A possibility for increasing the impact of integrity classes could thus be to intersperse them throughout the research experience, with small research integrity modules taking place at regular times, so that students can bring to mind the latest integrity issues they have encountered.

To further enhance the value of these regular research integrity seminars, we should also bear in mind the importance of making them attractive and social. Social comparison and emulation work best when people have the opportunity of imitating someone who is similar to them, who will thus be more likely to provide information that is relevant for their own personal integrity issues. Research integrity classes could thus be more attractive if taught by peers; that is, early-career researchers that are only slightly older than the students themselves. This early-career researchers' perspective is also important in the context of research integrity, since many research integrity issues involve Principal Investigator / Student conflict, for instance in cases involving power dynamics, such as cases involving moral or sexual harassment. Having a class taught by students that are on the same side of the Principal Investigator / Early-career researchers divide should be useful.

Making research integrity seminars attractive could involve discussing practical cases that have impacted the instructor's own career. This self-disclosure should give students a feeling that learning about these issues will have a major impact on their own life.

This principle of regular, Early Career Researchers-led seminars is already implemented in some highly successful settings, such as the ReproducibiliTEA community. ReproducibiliTEA started as an Early Career Researchers-led informal seminar in 2018, based on students' desire to discuss open science issues (Orben, 2019). The movement has had a huge success, leading to hundreds of local communities across the world.

At the high school and undergraduate levels, students are rarely involved in creating original research. In this context, promoting integrity often involves making sure that students' essays are not plagiarized, and show a proper respect for sources. Research has shown that students often have a fuzzy understanding of the boundary between plagiarism and appropriate quotation. Since integrity abuses increase in the context of uncertain rules, Erez Yoeli's three principles are crucial here. A simple way to clarify expectations could be to provide a checklist that the students have to sign before submitting an essay. Checklists have proved their efficiency in avoiding mistakes in a large range of domains, including engineering and medicine (Gawande, 2009). This checklist would include testifying that the students quoted all sources they drew from, and that they did not misrepresent what the sources said. Such a checklist would act by clarifying expectations, increasing observability (by showing that the teacher cares about these behaviours), and eliminating excuses (the students, by signing, testify that they have understood what is expected of them). It means that students cannot plagiarize without realizing that they are violating academic norms.

In general, nudges and behavioural interventions can have a maximal impact in conjunction with traditional incentives. The proper quotation of sources is a paradigmatic case of scientific integrity, as it shows both respect for past researchers and ensures that claims are backed by proper evidence. However, citing sources can be a source of cheating if students misrepresent past research to bolster dubious claims. A safe method to prevent the manipulation of sources consists in systematically checking the original sources, and sanctioning a paper if the sources are misquoted. Alternatively, the instructor could announce a random check of a few citations per paper, which would provide most of the benefit while saving time. Simply highlighting the possibility of random checks activates the feeling of being observed, and clarifies that proper citing of sources is a duty. Behavioural research has highlighted that increasing the possibility of an event from 0% to 1 or 2% can lead to a large change in behaviour, well beyond what would be justified by such a low shift in probability (Kahneman & Tversky, 1984).

Practical Case 2: Informal Academic Contexts

We can use nudges outside of formal educational contexts, in order to model the environment faced by students in order to improve their pro-social behaviour in their work as junior researchers.

For instance, the promotion of integrity among PhD students conducting experiments involves ensuring that they will conduct research in an open and transparent way. This involves posting data, research materials, and code online, in addition to preregistering one's experiments. These behaviours are somewhat costly since they can take a significant amount of researchers' time. However, it would be inefficient to make these practices mandatory, because there can be good reasons for not promoting them (for instance, the data can involve sensitive information, and should be kept private to protect participants' rights). We thus need soft incentives to promote them.

The simplest way of promoting open science practices is by turning them into the default. Institutional review board committees (IRBs) could for instance include a checklist that students would only have to cross if they plan to pre-register their experiments and put their data online. If students do not plan to use open science practices, IRBs could ask students to justify their decision. This can be described as a sludge, or the opposite of a nudge (Sunstein, 2020): adopting the non-optimal behaviour should be made more costly in terms of time spent justifying one's decisions.

Asking students when they submit an IRB application to commit to putting their data online is useful. However, it is not a perfectly safe procedure, since researchers often forget to put their data online. Universities should send simple reminders to students after the publication of their studies, asking them to put the data online if this has not already been done. Such reminders would have an impact both through the easy and social aspects of the EAST framework. These reminders are helpful, but also strongly communicate what is expected of researchers.

Moreover, universities could promote these behaviours by giving feedback to their students, sending them yearly reminders of how many pre-registrations they have done, how many studies they have made public, and comparing them to other students at their university. This social comparison aspect would signal that open science practices have become a new social norm.

More generally, promoting transparency could also generate other positive downstream effects. It is worth increasing observability whenever there is a high risk that some actor could cheat, discriminate, or take advantage of their position. Transparency offers the possibility to praise responsible conduct and to cause reputational damages to those who adopt free-riding or exploitative behaviour. Pro-integrity effects may be obtained in many academic domains, including administrative processes, selection procedures, workflows and evaluations within institutes, research groups or classes.

Ethical Nudging by Design

Nudges raise ethical issues, mostly autonomy concerns generated by the manipulative aspect these interventions. It is important to be aware and address those issues ahead of the application of a nudging intervention. Evaluation procedures, easily applicable by decision-makers, have been developed (Clavien, 2018; Lades & Delaney, 2022). They establish clear evaluation criteria that help to address the relevant ethical questions and to balance pro and contra reasons to apply a nudge. A systematic use of these procedures allows developing ethical nudges by design.

4. References

Allcott, H., & Rogers, T. (2014). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *American Economic Review*, *104*(10), 3003–3037. https://doi.org/10.1257/aer.104.10.3003

Anderson, M. S., & Adam, J. A. (2014). A Proposal for Considering Research Integrity from the Perspective of Behavioral Economics. *Journal of Microbiology & Biology Education*, *15*(2), 173–176. https://doi.org/10.1128/jmbe.v15i2.868

Bohner, G., & Schlüter, L. E. (2014). A Room with a Viewpoint Revisited: Descriptive Norms and Hotel Guests' Towel Reuse Behavior. *PLOS ONE*, *9*(8), e104086. https://doi.org/10.1371/journal.pone.0104086

Callaway, F., Hardy, M., & Griffiths, T. (2022). Optimal nudging for cognitively bounded agents: A framework for modeling, predicting, and controlling the effects of choice architectures. *Psyarxiv*. https://doi.org/10.31234/osf.io/7ahdc

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, aaf0918. https://doi.org/10.1126/science.aaf0918

Christensen, G., & Miguel, E. (2018). Transparency, Reproducibility, and the Credibility of Economics Research. *Journal of Economic Literature*, *56*(3), 920–980. https://doi.org/10.1257/jel.20171350

Clavien, C. (2018). Ethics of nudges: A general framework with a focus on shared preference justifications. *Journal of Moral Education*, *47*(3), 366–382. https://doi.org/10.1080/03057240.2017.1408577

DellaVigna, S., & Linos, E. (2022). RCTs to Scale: Comprehensive Evidence From Two Nudge Units. *Econometrica*, *90*(1), 81–116. https://doi.org/10.3982/ECTA18709

Dimant, E., van Kleef, G. A., & Shalvi, S. (2020). Requiem for a Nudge: Framing effects in nudging honesty. *Journal of Economic Behavior & Organization*, *172*, 247–266. https://doi.org/10.1016/j.jebo.2020.02.015

Douglas, H. (2014). Scientific Integrity in a Politicized World. In *Logic, Methodology and Philosophy of Science.* (pp. 253–268).

Errington, T. M., Mathur, M., Soderberg, C. K., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Investigating the replicability of preclinical cancer biology. *ELife*, *10*, e71601. https://doi.org/10.7554/eLife.71601

Fishbane, A., Ouss, A., & Shah, A. K. (2020). Behavioral nudges reduce failure to appear for court. *Science*. https://doi.org/10.1126/science.abb6591

Gawande, A. (2009). *The Checklist Manifesto: How to Get Things Right* (First edition). Metropolitan Books.

Halpern, D. (2016). *Inside the Nudge Unit: How Small Changes Can Make a Big Difference* (Reprint edition). WH Allen.

Hansen, P. G., Schilling, M., & Malthesen, M. S. (2021). Nudging healthy and sustainable food choices: Three randomized controlled field experiments using a vegetarian lunch-default as a normative signal. *Journal of Public Health*, *43*(2), 392–397. https://doi.org/10.1093/pubmed/fdz154

Integrity Games. (2022). https://integgame.eu/

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*(4), 341–350. https://doi.org/10.1037/0003-066X.39.4.341

Lades, L. K., & Delaney, L. (2022). Nudge FORGOOD. *Behavioural Public Policy*, *6*(1), 75–94. https://doi.org/10.1017/bpp.2019.53

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology*, *73*(1), annurev-psych-020821-114157. https://doi.org/10.1146/annurev-psych-020821-114157

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Orben, A. (2019). A journal club to fix science. *Nature*, *573*(7775), 465–465. https://doi.org/10.1038/d41586-019-02842-8

Silva, A., & John, P. (2017). Social norms don't always work: An experiment to encourage more efficient fees collection for students. *PLOS ONE*, *12*(5), e0177354. https://doi.org/10.1371/journal.pone.0177354

Stricker, J., & Günther, A. (2019). Scientific Misconduct in Psychology. *Zeitschrift Für Psychologie*, *227*(1), 53–63. https://doi.org/10.1027/2151-2604/a000356

Sunstein, C. R. (2014). Nudging: A Very Short Guide. *Journal of Consumer Policy*, *37*(4), 583–588. https://doi.org/10.1007/s10603-014-9273-1

Sunstein, C. R. (2020). Sludge Audits. *Behavioural Public Policy*, 1–20. https://doi.org/10.1017/bpp.2019.32

Thaler, R. H., & Sunstein, C. R. (2021). *Nudge* (The Final edition). Yale University Press.

Yoeli, E. (2018). *How to motivate people to do good for others*. https://www.ted.com/talks/erez_yoeli_how_to_motivate_people_to_do_good_for_others