Is Convergent Realism an Appropriate Method for Evaluating Ethics?

Muki Haklay, University College London, m.haklay@ucl.ac.uk

––––––––––––––––––

This article serves an invited response to Chris Santos-Lang's "The Method of Convergent Realism" (2022). I respond from a social constructionist position, mixed with critical realism, that challenges some of the premises of Santos-Lang's paper. I question if the use of convergent realism is appropriate for the example that is at the centre of the paper—the development of procedures and practices of an independent, community-led ethics review board. I also question the utility and risks that are associated with another central suggestion in the paper—the use of automated machine-learning and natural language processing system to support ethical screening and review.

I write this response from the following position: I am neither a philosopher, nor an ethicist. However, I have an interest in ethical conduct within my own research practice, and I have constructively engaged with the Research Ethics Committee in my institute both to examine my own research, and to consider how ethical practices are integrated within postgraduate research activities. I was also involved in the academic team that developed ethics training within an engineering department at UCL (Doffour et al. 2010). Finally, I am familiar with bottom-up community research projects and the wider areas of citizen science and crowdsourcing which are relevant to this discussion. With this explanation in place, let us turn to Santos-Lang's paper.

Experimentation and Ethics Review

Santos-Lang's interesting and thought-provoking piece links a form of the scientific process, common within parts of the natural sciences, with ethics review for research proposals. Assuming a clear true answer for the "right" ethical procedures for a given research project exists, it is suggested that there is potential for supporting ethical procedures through machine-learning and natural language processing, demonstrated with a hypothetical ability of the IBM Watson system. Santos-Lang suggests that we adopt a method of convergent realism in addressing different problems including the ethical scrutiny of proposed research activities.

Santos-Lang proposes three steps for operationalising convergent realism, namely: articulation of reasoning; independent testing of reasoning; and enforcing retesting and re-validation. To enable the enforcement, Santos-Lang describes five core mechanisms: transparency, conflict resolution, amendment process, expiration dates, and objectivity.

Santos-Lang's approach seems relevant for natural science research projects, especially projects that rely on open science and large-scale public participation. In such cases, if we assume that there is a single truth, such as the classification of a given galaxy in the observable universe as elliptical or not (Lintott et al. 2008), then it is possible to use the three-stage procedure to reach this truth. The procedure and its mechanisms will also work very well in medical research such as patients-led research. For example, for the Amyotrophic Lateral Sclerosis (ALS) patients who have read about a small-scale experiment with lithium carbonate as a potential medicine that slows the rate of the disease (Wicks 2018). They organised to carry out a controlled study through the platform PatientsLikeMe

(https://www.patientslikeme.com/) and discovered that their independent testing showed that the medicine doesn't work, thus making progress towards the truth. Yet, the article demonstrates, in my opinion, a severe limit of convergent realism in addressing social science and humanities problems. I will argue that handling and judging ethical review application is such a problem.

In what follows, I briefly discuss the multiple purposes of ethics process, as they are practiced in institutional Research Ethics Committees (RECs, also known as Institutional Review Board or IRB). I will then take up the problem in the assumption of truth within ethics applications and reviews, and then look at the challenges that the Belleville Research Ethics Committee (BREC) face because of the way it operates. Finally, I will explore the potentials and limitations of automating such processes.

The History and Practice of Ethics

While ethics is a clear field of inquiry within western philosophy, and provides a way to consider right and wrong behaviour in a human society (Driver 2022), the role of RECs and ethical processes, in general, resides within the area of applied ethics. Moreover, one can argue that while applied ethics has a place in philosophy, what the REC is doing is practised ethics, which changes dynamically when people bring ethics to the concrete evaluation and approval of proposed human activities. Here, we need to understand the development of ethics within a specific area of human practice, and then consider what these circumstances reveal to us about the aims and objectives of ethical reviews.

As Mitcham (2009) shows regarding the field of engineering, the development of ethics within a specific area of practice reveals multiple goals. Codes of ethics go through different historical stages—first, informal concepts of ethics as the profession of engineers formalised and identified commonalities in practices. Second, ethics evolves as a form of loyalty towards the profession, the employer, and the client. For example, such codes may include guidance not to undercut another member of the profession in terms of pricing. Third, ethics alters for the purposes of the profession. In the case of engineers, for example, such a purpose may consist of emphasising efficiency. Fourth, comes a commitment to public safety, health, and welfare. Fifth, and recently, arises care for the environment and sustainability. Davis (1998) makes a similar case in describing engineering codes of ethics. It is unsurprising that the resulting ethical codes are simultaneously (and sometimes contradictorily) try to protect the standing of the profession, protect the employer, consider the needs of the client, try to ensure public safety and welfare, and care for the environment.

What we need to take from this work is that ethics codes continuously evolve, are an intensely social process, and need continuous interpretation in order to make them operational.

What is also obvious, to anyone involved in active scientific research, is the active evolution of the role of RECs. I recall the first time that I heard about IRBs at a conference in the early 2000s. American colleagues, who were carrying out research on user interaction with

computerised maps, discussed in detail how to develop an IRB application and the complexities of planning for it. At the time I was puzzled, since in the UK I had not heard of anyone asking for ethics approval for such experiments. Today, usability testing in the UK goes through the REC routinely. Even so, differences among disciplines remain and while some colleagues at UCL will secure an ethical approval before carrying out interviews, others will assume that their research is exempt from approval.

Ethics and Truth

While the previous section shows that ethics is always "work in progress", and so the practice of the REC must adapt and evolve, we should ask about the aspect of truth (in the sense that Santos-Lang uses it) for these processes. At any given time, a set of codified norms will serve as the basis for an ethics evaluation. For example, in the areas that are under the jurisdiction of the European Union, 2018 marked the coming into force of the General Data Protection Regulation (GDPR). As Clarke et al. (2019) discuss about clinical research in Ireland, these new regulations have an impact of the way research can be conducted and require an adjustment by researchers and patients to enable research to continue, including adjusting the process of consent to enable such research. The regulations and their interpretation are therefore an example of codified norms. These are, despite their continual evolution, a basis for assessment and include disciplinary codes of ethics, guidance from research funders, and institutional procedures that define codes of practice and regulations. Issues such as the insurance policy of the institution lend additional context for what its researchers may or may not be allowed to do. Moreover, and this is a common practice within RECs, a set of project archetypes exists for which the issues are well understood and the scrutiny is basic. Unfortunately, the evidence that exist points to inconsistency between RECs in evaluation of similar research applications (Angell et al. 2006; Angell et al. 2007; Trace 2017).

In this sense, one can argue that there is a potential "current truth" or a "temporary truth" for the cases that are easier to decide, to which the method that is offered by Santos-Lang can aim. For the sake of assessing the relevance of convergent realism for ethics, we can focus on these cases and examples in the BREC procedures that help us reach it.

BREC Procedures and "Crowdsourcing" Ethics Evaluation

While unfortunately the full context of BREC is not provided in Santos-Lang, for the sake of this response, I will suggest that it is as follows.

A BREC-like committee is composed of individuals who volunteer to act as members of the committee. The committee is not associated with any institution; rather, it deliberately operats within the realm of open science and citizen science. That is, the committee comes together to evaluate an application on an ad-hoc basis and likely without financial compensation from the applicant (unlike private IRBs which exist in the USA). As a result, and to avoid burden on the volunteers, the number of applications that are evaluated is limited. This constraint makes the BREC-like committee somewhat like a jury. The number

of such bodies is only limited by the number of individuals who will be volunteering to join and form it.

At first look, this approach seems like a good idea. In the case of classifying galaxes, Watson and Floridi (2018) show that epistemologically crowdsourcing yields highly robust results—arguably more than what is possible within professional research due to resource limitations. Alas, I would not expect such convergence and approach towards the truth by these BREC-like committees. The body of knowledge necessary to make a judgement on ethics applications is quite significant and, so, permanent staff work in universities that monitor and facilitate ethics committees. There is no doubt in my mind that people with the necessary skills and experience exist in the world outside research institutions—be they lawyers, retired researchers, religious leaders, patients, and community activists, and so on. Yet, apart from BREC, there are no other examples of spontaneous research committees. There is evidence from citizen science projects that participants are concerned about their data exactly because of their lack of expertise and that they might get things wrong (Eveleigh et al. 2014). We can expect reluctance and concern by lay people to make judgements on research ethics that include two areas from which the general public is regularly excluded: law and science. We can see some of this struggle in the experience of lay members within existing IRBs (Dyer 2004). Thus, creating of BREC-like committees faces major barriers.

Even if we assume the proliferation of such committees, it remains very difficult to reach a "temporary truth" through crowdsourcing. As noted above, institutional committees are inconsistent. Without the institutional boundaries, each BREC-like committee will rely on its members values, worldviews, and understanding of the research at hand. Disagreements are likely which will undermine this potential truth. If my reading is correct, such circumstances suggest a role for automated knowledge-based systems. Since, as noted, part of the basis for IRBs are codified norms, these norms and a body of existing applications and their associated corrections and approvals can be used as a basis for such a system. In the final part of this response, I turn to this question.

Can IBM Watson Guide Ethics Procedures?

The use of a machine with a capacity for natural language processing, information retrieval, and automated reasoning, depends on the nature of the corpus of documents that are the basis of the responses. Since, as noted, part of the ethical screening is based on existing laws and regulations, those can be captured and assessed against the description of a corpus of other ethics applications that were evaluated and accepted. The codes of ethics can include disciplinary codes of ethics, laws and data protection regulations. As is common in cases where computers are used, the forms that are submitted for such an evaluation can be adjusted to make the form more machine readable. The beauty of systems such as Watson is that the effort by humans to prepare the data for the algorithms is reduced.

While I failed to find current literature on automating research ethics applications, it is clear that many parts of the process can benefit from the current abilities of automated systems (so called "Artificial Intelligence"). These systems can surely evaluate the readability and

clarity of the consent form and information sheet, compare the procedures that the researchers describe to a recent corpus of similar research, and so on. Yet, risks remain. The research might be similar to another project, but different in a subtle way. For example, by changing the target audience from "students at the university" to "women in the neighbourhood next to the university" (which happens to be deprived and stigmatised), the meaning of the research and its implications can change dramatically. Even the cartographic usability experiment that I described above will change. However, this work can be supported by researchers indicating changes on the form in a machine-readable way.

Assuming the existence of such a system, it is likely that it can, as Santos-Lang suggests, provide support to BREC-like committee members. Yet, a significant issue resides with the ability of such a system to clarify reasoning, articulate it in an accessible way to lay people, and present the information in such a way that it can assist the decision. The current problems in the use of such systems in public administration—which include biases, opacity, and risks to personal information (Sobrino-García 2019)—indicate that there is a long way to go for this to happen. These systems are not developed enough to provide such counselling services.

Summary

Santos-Lang offers a method for research and suggests that it can support the process of ethics evaluation through a form of crowdsourcing and provision of support and procedures for such research ethics committees. This response raises challenges to his suggestion. First and foremost, I argued that there is no clear truth towards which the activities of these multiple RECs can progress. Second, I highlighted the complexity of operating a REC outside institutional boundaries. Third, I questioned the possibility of knowledge support system that will enable such RECs to operate.

I do not dismiss or undermine the importance of the BREC procedures and the growth in ethical standards within bottom-up research. For different communities of practice to coalesce around different types of community-led research, the BREC procedure can provide guidance—not through an algorithmic process, but through detailed deliberation. There are already such practices; for example, within the Do It Yourself (DIY) Biology community (Eggleson 2014), or within the Public Laboratory of Open Technology and Science". For such groups, potential exists for developing the ideas forwarded by Santos-Lang.

References

Angell, Emma L., Clare J. Jackson, Richard E. Ashcroft, Alan Bryman, Kate Windridge, Mary Dixon-Woods. 2007. "Is 'Inconsistency' in Research Ethics Committee Decision-Making Really a Problem? An Empirical Investigation and Reflection." *Clinical Ethics* 2 (2): 92-99.

Angell, Emma L., Alex J. Sutton, Kate Windridge, Mary Dixon-Woods. 2006. "Consistency in Decision Making by Research Ethics Committees: A Controlled Comparison." *Journal of Medical Ethics* 32 (11): 662-664.

Clarke, Niamh, Gillian Vale, Emer P. Reeves, Mary Kirwan, David Smith, Michael Farrell, Gerard Hurl, and Noel G. McElvaney. 2019. "GDPR: An Impediment to Research?" *Irish Journal of Medical Science (1971-)* 188 (4): 1129-1135.

Davis, Michael. 1998. *Thinking Like an Engineer: Studies in the Ethics of a Profession.* Oxford University Press.

Dyer, Sarah. 2004. "Rationalising Public Participation in the Health Service: The Case of Research Ethics Committees." *Health & Place* 10 (4): 339-348.

Driver, Julia. 2022. "Moral Theory." *The Stanford Encyclopedia of Philosophy*. Fall Edition. Edited by Edward N. Zalta. https://plato.stanford.edu/archives/fall2022/entries/moral-theory/.

Duffour, Philippe, Sarah Bell, Christian Solberg, Muki Haklay. 2010. "Reflection of UCL Civil and Environmental Engineering Experience in Integrating Ethics in Engineering Education." https://www.researchgate.net/publication/228852281 _Reflection_of_UCL_Civil_and_Environmental_Engineering_experience_in_integra ting_ethics_in_engineering_education.

Eggleson, Kathleen. 2014. "Transatlantic Divergences in Citizen Science Ethics— Comparative Analysis of the Diybio Code of Ethics Drafts of 2011." *Nanoethics* 8 (2): 187-192.

Eveleigh, Alexandra, Charlene Jennett, Ann Blandford, Philip Brohan, Anna L. Cox. 2014. "Designing for Dabblers and Deterring Drop-Outs in Citizen Science." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2985-2994. doi: 10.1145/2556288.2557262.

Lintott, Chris J., Kevin Schawinski, Anže Slosar, Kate Land, Steven Bamford, Daniel Thomas, M. Jordan Raddick et al. 2008. "Galaxy Zoo: Morphologies Derived from Visual Inspection of Galaxies from the Sloan Digital Sky Survey." *Monthly Notices of the Royal Astronomical Society* 389 (3): 1179-1189.

Mitcham, Carl. "A historico-ethical perspective on engineering education: From use and convenience to policy engagement." *Engineering studies* 1, no. 1 (2009): 35-53.

Puiblic Lab "Ethics and Politics" Online https://publiclab.org/wiki/ethics-and-politics (accessed 31 July 2022)

Santos-Lang, Chris. 2022. "The Method of Convergent Realism." *Social Epistemology Review and Reply Collective* 11 (1): 33-49. https://wp.me/p1Bfg0-6t5

Sobrino-García, Itziar. 2021. "Artificial Intelligence Risks and Challenges in the Spanish Public Administration: An Exploratory Analysis Through Expert Judgements." *Administrative Sciences* 11 (3): 102. doi: 10.3390/admsci11030102.

Trace, Samantha, and Simon Erik Kolstoe. 2017. "Measuring Inconsistency in Research Ethics Committee Review." *BMC Medical Ethics* 18 (1): 1-10.

Wicks, Paul. 2018. "Patient, Study Thyself." 2018. *BMC Medicine* 16 (1): 1-3.