A First Survey on an Atlas of Intelligence

Sankalp Bhatnagar[1,2], Anna Alexandrova[1,4], Shahar Avin[3], Stephen Cave[1], Lucy Cheke[1,11], Matthew Crosby[1,7], Jan Feyereisl[8,9], Marta Halina[1,4], Bao Sheng Loe[5], Seán Ó hÉigeartaigh[1,3], Fernando Martínez-Plumed[6], Huw Price[1,3], Henry Shevlin[1], Adrian Weller[1,12], Alan Winfield[1,10], and José Hernández-Orallo[1,6,13]

[1] Leverhulme Centre for the Future of Intelligence, UK.
[2] The New School, New York.
[3] Centre for the study of Existential Risk, University of Cambridge, UK.
[4] Department of History and Philosophy of Science, University of Cambridge, UK.
[5] Psychometrics Centre, University of Cambridge, UK.
[6] Universitat Politècnica de València, Spain.
[7] Imperial College, UK.
[8] AI Roadmap Institute, Czech Republic.
[9] GoodAI, Czech Republic.
[10] Bristol Robotics Laboratory, UWE Bristol, UK.
[11] Department of Psychology, University of Cambridge, UK.
[12] Alan Turing Institute, UK.
[13] Corresponding Author: jorallo@dsic.upv.es.

Abstract. In this note, we include the details of a questionnaire we designed in the summer of 2017 and the results of the survey we conducted on experts from different communities: artificial intelligence, animal cognition, psychology, philosophy, design and some others. We analyse the results and the feedback we received.

1 Introduction

AI research is blooming, fuelled by a range of algorithmic and computational developments in machine learning, robotics and computational cognitive modelling. However, the actual pace and direction of AI progress are largely unassessed and hard to extrapolate or direct. The main reason is that we lack tools to properly evaluate, compare and classify AI systems, and thus guide the future of the field. The comparison of AI systems with human and non-human intelligence is done in an informal and subjective way, usually with contradicting assessments, especially when looked at in hindsight (Hayles, 1996; Brooks, 1997; Pfeifer, 2001; Shah et al, 2016).

This requires a novel platform, an 'atlas of intelligence', that integrates an extensive inventory of cognitive systems, a behavioural test catalogue (with test batteries that could be aggregated into dimensions) and an experimentation repository (the results from actual measurements). The platform would be populated in a collaborative fashion, facilitating cross-comparison and reproducibility (Aarts et al, 2015; Vanschoren et al, 2015) and testing across a wide range

of individuals (Hernández-Orallo, 2017). The atlas would represent a new cartographic endeavour for a better understanding of the geography of the space of intelligence, which is changing at both the ontological and epistemological levels.

More details about the atlas can be found in a companion paper (Bhatnagar et al, 2018).

2 Feedback

Apart from the scientific questions needed to build such a platform, its success depends on the engagement of the (research) community and other stakeholders. It is crucial then to identify whether the needs, dimensions and elements represented are well aligned with the potential users and contributors. Consequently, we conducted a preliminary survey to get feedback from researchers and other potential users in many different areas. We targeted different communities: artificial intelligence, animal cognition, psychology, philosophy, design and some others. In the end, we got 33 responses, whose averages are shown in Fig. 1.

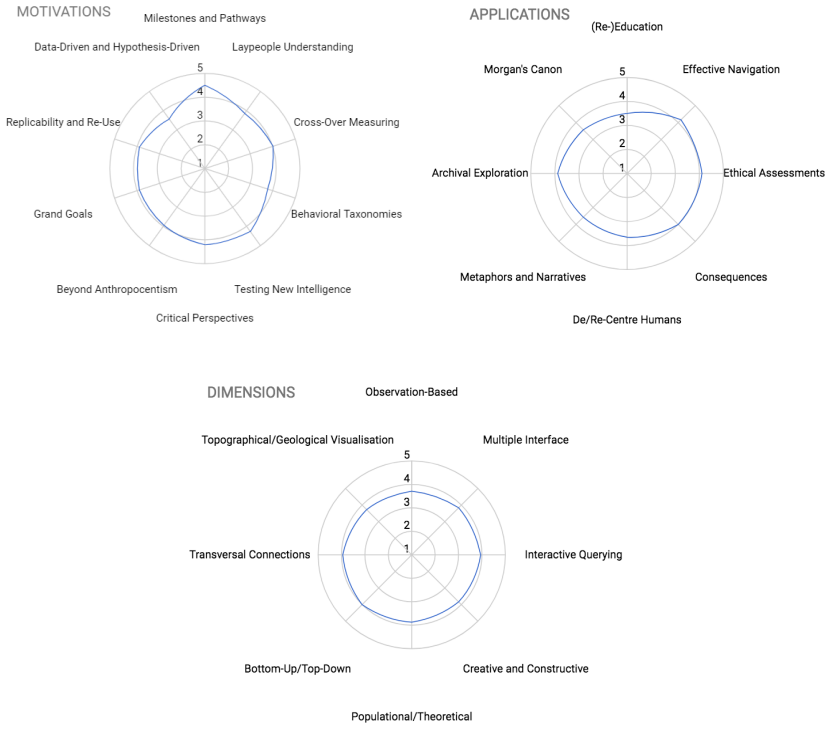

Fig. 1: Average survey results.

For the kinds of intelligence, all questions had at least 66% of positive assessments, and are not represented in the figure. Overall, the responses were mostly

in agreement with our conceptions. This is of course a sign that what we proposed was well received, but also that the questionnaire was not very informative and not very useful to determine those choices that are not well aligned with users and contributors. Even if the average results were good for all items, we found a few items for which there was a relatively high number of disagreements. In particular:

- **Grand Goals**: There was a 9.1% of completely disagreement. Some respondent found this ambiguous (the questionnaire only included the descriptions found in Tables 1 to 4), or expressions such as inspire the next generation were considered too "grand".
- **Data-Driven and Hypothesis-Driven**: There was a 12.1% of complete disagreement. It seems that some people think that the formulation of the question is too negative by saying that some disciplines are not up-to-date in the way they are making science.
- **Morgans Canon**: There was a 9.1% of completely disagreement. Some people might think that there is no need for a re-understanding in animal cognition while some people in AI may not know what Morgans Canon is.

More interesting were the open comments. Next we discuss some of them and possible answers.

We had some comments around the need of considering the wisdom of the crowd and integration, or how we could integrate "collective intelligence" here, i.e., what a community of agents is able to do with different topologies and communications. Even if we mentioned collective intelligence, some respondents wanted this to be more explicit and elaborated.

Some other experts highlighted that we should not focus solely on behavioural features, but also some other dimensions such as processing (for computers) and brain sizes/organisation ("neural substrates" for natural systems). In other words, the atlas should cover "black-box" behaviour but should also open the box on some dimensions, especially those that can be compared across a range of systems. This was the intention, as the transversal connection shows, but not explicit on other items.

We also received suggestions about whether the atlas should be theory-free. This was a very important question that presented itself during the internal discussions: how theory-laden is the atlas meant to be? From the two extremes (theory-free and bound to a particular theory, or reduced group of theories), we take these comments by saying that the atlas should be theory-agnostic or theory-plural (we already used the term agnostic in the questionnaire). This links to a related comment from another expert that reminded us that observations (especially if the atlas is based on evidence) are usually driven by theories (or hypotheses). Consequently, we have to be very careful about the traceability of the data through appropriate metadata, and what theory lies beneath, especially if we want to cross data with different theories and get consistent results and representations.

Similarly, some experts raised the choice between many independent "intelligences" (or just infinitely many task performances) and a unified (or even

monolithic) intelligence. We do not share any of these two extremes, but consider that there is some structure in the space of intelligence. Nonetheless, the atlas must not constrain to any precise structure of intelligence (hierarchical, multidimensional with overlaps between the dimensions, or others), but should allow users to define their structures when combining data into more aggregate dimensions and ultimately visualisations.

There was a suggestion to include both qualitative and quantitative dimensions, which we agree. Even if maps are usually represented along quantitative dimensions (longitudes and latitudes), some elements of the maps are qualitative (water vs land) and so are represented in these maps. There are some dimensions that, because of our lack of detail, or because they are intrinsically discrete, it is better to represent them as qualitative dimensions, using quadrants or other kinds of representations.

Finally, there was a suggestion to have historical information about technology, i.e., to be able to compare how different kinds of architectures show different trajectories in time in terms of some dimensions. This comment goes beyond our reference in the questionnaire to past, present and future information, and asks for a richer coverage of these issues, which again is related to careful handling of metadata, which must include time.

Overall, we think the form and the series of direct interviews with experts in other fields helped us to refine and validate the specifications, at least in the way that all the recognised possibilities were desirable, at least in principle. However, as this is a very ambitious project, we must also need to recognize which motivations and applications are feasible in the near future and which ones could have more impact. This would help us set the priorities, once the initiative moves to design and planning stages. Consequently, as future work, we are considering the possibility of a wider poll at some conferences, with a more precise specification as a reference, where we ask for priorities and feasibilities (possibly asking to rank the motivations and applications instead of assessing them individually).

Appendix

In this appendix we include the items and their descriptions, as they appeared in the questionnaire. Tables 1, 2, 3, 4 show the motivations, applications, dimensional manipulations and entities, respectively.

Item	Description (as appeared in the questionnaire)
MILESTONES AND PATHWAYS	"Need to map out the milestones of where human, AI and hybrid intelligence is going (also historically) and spot opportunities and risks."
LAYPEOPLE UNDERSTANDING	"Difficult access for lay people to complexity contours of intelligence (e.g., "the map that came to life")."
CROSSOVER MEASURING	"Lack of a more systematic crossover test measuring of different species and technologies."
BEHAVIORAL TAXONOMIES	"Need for taxonomies and representations at the behavioral level, in contrast to phenotypic, ethological, genotypic or neurological approaches (e.g., the "Allen Brain Atlas")."
TESTING NEW INTELLIGENCE	"Yet unsolved but more urging question of how to develop tests beyond the Turing test and for new kinds of intelligence?"
CRITICAL PERSPECTIVE	"Insufficiently critical and well-founded understanding of intelligence for society, researchers and policymakers."
BEYOND ANTHROPOCENTRISM	"Still too anthropocentric view of intelligence."
GRAND GOALS	"Inspire the next generation (a compass for research, a reference for narratives, etc.)."
REPLICABILITY AND REUSE	"Poor replicability and re-use when collating the state of the art from the same and especially from different disciplines."
DATA-DRIVEN AND HYPOTHESIS-DRIVEN	"The new way of making science (e.g., through a collaborative data science approach) hasnt reached some disciplines (e.g., animal cognition, AI,)."

Table 1: Motivations for the atlas, as they were identified in August 2017 and included (verbatim) in the questionnaire.

Acknowledgements

The initiative was supported by the Leverhulme Trust via the CFI. J. H-Orallo and F. M-Plumed were supported by EU (FEDER) and the Spanish MINECO under grant TIN 2015-69175-C4-1-R and by GVA under grant PROMETEOII/2015/013 and by the Air Force Office of Scientific Research under award number FA9550-17-1-0287. J. H-Orallo also received a Salvador de Madariaga grant (PRX17/00467) from the Spanish MECD for a research stay at the Leverhulme Centre for the Future of Intelligence (CFI), Cambridge, and a BEST grant (BEST/2017/045) from the GVA for another research stay also at the CFI. A. Weller acknowledges support from the David MacKay Newton research fellowship at Darwin College, the Alan Turing Institute under EPSRC grant EP/N510129/1 & TU/B/000074, and the Leverhulme Trust via the CFI.

Item	Description (as appeared in the questionnaire)
(RE-)EDUCATION	"Animals and robots presented in an integrated, comparable way, at school, universities, research and governments, much like children view animal and robot personas today (e.g., Pokemon cards)."
EFFECTIVE NAVIGATION	"Faster/safer routing in the intelligence space routes to AI milestones and human enhancement, danger areas (risks, things we need most, gaps)."
ETHICAL ASSESSMENTS	"Assess areas of possible ethical concern (animal/robot suffering, human suffering, uncanny valleys, . . .)."
CONSEQUENCES	"Analyze (especially in advance) the population of creatures affected by research, law, environment, technology, etc., in a critical way (e.g., islands and gaps not covered by certain laws, ideologies or stances, such as veganism)."
DE/RE-CENTRE HUMANS	"Be able to locate our species in the atlas, groups of people and even individuals in terms of different dimensions. Raising awareness of "glocality""
METAPHORS AND NARRATIVES	"Allow varied animal metaphors and narratives be used across domains in new and more systematic ways."
ARCHIVAL EXPLORATION	"Query the historical progress of intelligence on Earth (evolutionary, social, technological), including extinct animals and extinct computer/AI systems. Address cognitive/behavioral changes in populations such as the Flynn effect, Google effect, animal domestication, etc."
MORGANS CANON	"Re-understand and apply Morgans canon more generally (how to quantify more/less complex for animals but also AI)."
REPLICABILITY AND REUSE	"Poor replicability and re-use when collating the state of the art from the same and especially from different disciplines."
UNIFICATION	"Define unified definitions of intelligence and other behavioral features. Define universal tests."

Table 2: Potential applications for the atlas, as they were identified in August 2017 and included (verbatim) in the questionnaire.

Item	Description (as appeared in the questionnaire)
OBSERVATION - BASED	"Agnostic to particular theories and hypotheses, minimizing assumptions (the users will do their hypotheses/theories from it)."
MULTIPLE INTERFACE	"Projections/aggregations of data into maps and other representations: atlas is collection of these maps."
INTERACTIVE QUERYING	"Different methods of interrogating the atlas (queries, filters, . . .)."
CREATIVE AND CONSTRUCTIVE	"Creating new features/territories of intelligence by interrogating/querying maps across species and technologies, combining particular behaviors."
POPULATIONAL/THEORETICAL	"Include population analyses (existing individuals/groups) vs. theoretical analyses (possible individuals/groups)."
BOTTOM-UP/TOP-DOWN	"Distinguish gross behavioral capacities from underlying psychological mechanism at different levels of aggregation (specific vs general)."
TRANSVERSAL CONNECTIONS	"Allow connections between dimensions and capacities (not orthogonal, relational), also attending to non-behavior traits (evolutionary, physical, etc.)."
TOPOGRAPHICAL/GEOGRAPHICAL VISUALIZATION	"Using colors, contours, textures and not only cartesian representations."

Table 3: Potential dimensional manipulations for the atlas, as they were identified in August 2017 and included (verbatim) in the questionnaire.

Item (as appeared in the questionnaire)
"General and narrow"
"Individual and collective (groups, teams, species,)"
"Biological and artificial"
"Hybrid (extended/enhanced minds, cyborgs)"
"Novel and old (existing, extinct)"
"Distributed and centralized"
"Alien and fictional (biological and theoretical)"

Table 4: Types of potential entities to be considered for the atlas, as they were identified in August 2017 and included (verbatim) in the questionnaire.

Bibliography

Aarts A, Anderson J, Anderson C, Attridge P, Attwood A, Fedor A (2015) Estimating the reproducibility of psychological science. Science 349(6251):1–8

Bhatnagar S, Alexandrova A, Avin S, Cave S, Cheke L, Crosby M, Feyereisl J, Halina M, Loe BS, Ó hÉigeartaigh S, Martínez-Plumed F, Price H, Shevlin H, Weller A, Winfield A, Hernández-Orallo J (2018) Mapping Intelligence: Requirements and Possibilities

Brooks RA (1997) From earwigs to humans. Robotics and Autonomous Systems 20(2-4):291–304

Hayles NK (1996) Narratives of artificial life. Future Natural: Nature, Science, Culture London: Routledge pp 146–64

Hernández-Orallo J (2017) The measure of all minds: evaluating natural and artificial intelligence. Cambridge University Press

Pfeifer R (2001) Embodied artificial intelligence 10 years back, 10 years forward. In: Informatics, Springer, pp 294–310

Shah H, Warwick K, Vallverdú J, Wu D (2016) Can machines talk? comparison of eliza with modern dialogue systems. Computers in Human Behavior 58:278–295

Vanschoren J, Bischl B, Hutter F, Sebag M, Kegl B, Schmid M, Napolitano G, Wolstencroft K, Williams AR, Lawrence N (2015) Towards a data science collaboratory. Lecture Notes in Computer Science (IDA 2015) 9385