Notes on Measures for Information Access in Neuroscience and AI Systems

Giulio Prevedello[1,2,*,a], Luis Marcos-Vidal[3,*,b], Emanuele Brugnoli[2,4,c] and Pietro Gravino[1,2,d]

[1]*Sony Computer Science Laboratories Paris, Paris, France*
[2]*Centro Ricerche Enrico Fermi, Rome, Italy*
[3]*Hospital del Mar Research Institute, Barcelona, Spain*
[4]*Sony Computer Science Laboratories Rome - Joint Initiative CREF-SONY, Rome, Italy*

Keywords: Awareness, Information Access, Structural Connectivity, Functional Connectivity, Operational Connectivity, Intrinsic Ignition, Conditional Mutual Information.

Abstract: With the increasing prevalence of artificial systems in society, it is imperative to ensure transparency in machine decision processes. To better elucidate their decisions, artificial systems must possess an awareness of the information they handle. This includes the understanding of information flow, integration and impact on the final outcome. A specific facet of awareness, termed access-consciousness, denotes the ability of information to be utilised in reasoning and the rational control of action (and speech). This study proposes a method for measuring access to information within a system by examining the communication dynamics among its components, specifically focusing on connectivity. To achieve this, we initially delineate the various types of connectivity in the brain and then propose their translation to artificial systems. Structural connections are highlighted as mechanisms enabling one component to access information from another. Additionally, we explore functional connectivity, which gauges the extent to which information from one component is utilised by another. Finally, operational connectivity is introduced to describe how information propagates from one component to the entire system. This framework aims to contribute to a clearer understanding of information access in both biological and artificial systems.

1 INTRODUCTION

One of the objectives of Artificial Intelligence (AI) research is to create systems that can adapt their actions to the human user. In parallel, many efforts are spent to make the machine's decision process more understandable and transparent to instil confidence and facilitate interactions. This is reflected by the numerous works about human-centred AI, AI alignment or explainable AI in the scientific community (Koster et al., 2022). At the same time, public attention invested this field after the recent achievements of large language models (Min et al., 2023).

Explainability in humans is related to being "aware of something". But what does awareness mean? It is true that both awareness and consciousness have been mongrel terms that encompass multiple phenomena (or at least multiple parts of the same phenomenon), but when referring to AI and explainability, we can narrow their definition. Following the famous distinction made by Ned Block between phenomenal-consciousness and access-consciousness (Block, 1995), in this work we will only focus on the latter. According to Block's definition, information is access-conscious if it can be used for reasoning, and it is poised for rational control of action and of speech (Block, 1995). It is then only related to the content of the information and who is able to use it (Chalmers, 1997, "easy problem"), in contrast to phenomenal consciousness, which is related to subjective experience (Chalmers, 1997, "hard problem").

Access-consciousness, whether in humans or artificial systems, requires the acquisition of external information, internal processing, and the transmission of this processed information to the system's actua-

[a] https://orcid.org/0000-0002-9857-2351
[b] https://orcid.org/0000-0003-2135-1478
[c] https://orcid.org/0000-0002-5342-3184
[d] https://orcid.org/0000-0002-0937-8830
[*]Equally contributed

Prevedello, G., Marcos-Vidal, L., Brugnoli, E. and Gravino, P.
Notes on Measures for Information Access in Neuroscience and AI Systems.
DOI: 10.5220/0012596000003636

In *Proceedings of the 16th International Conference on Agents and Artificial Intelligence (ICAART 2024) - Volume 3*, pages 1429-1435
ISBN: 978-989-758-680-4; ISSN: 2184-433X

tors. These actuators then generate an output, such as an action or decision, which is external to the system. Therefore, it is essential to comprehend how information flows within the system's components and how access to information occurs. This study specifically examines the information propagation in both biological and artificial systems, making comparisons between the brain and AI models.

For both domains we assume any system is made of components that receive, process and send information to others. If information can travel from one component to another, a link is established between them. Thus, such a system can be represented as a graph, whose structural connectivity traces where information can flow between its nodes. No particular restrictions are imposed to the possible topology of connections. For example, a sender node could provide information to several others and a receiver nodes could get signals from many sources. In general, structures such as loops, trees, disconnected components are acceptable, and it is possible that the system's structural connectivity can change over time.

While one component can send information to another, if there is a structural connection, it is possible that no information is provided or that the information is redundant with respect to the signals from other sources. Thus, we define the functional connectivity of the system the ensemble of links where the flow of information is useful, in the sense that it is (functionally) used by the receiver and integrated in its output.

Of note, a flow of information presuppose a stream of data that are continuously passed between components. Yet, it is possible to restrict the notion of connectivity further to the path of components activation triggered by a single datum as it is elaborated by the system. The ensemble of these paths, over several data, defines the operational connectivity.

In the following, we delve into these three concepts of connectivity (structural, functional and operational) and, based on those, we review possible measure for access of information, tracing parallel between neurological and AI systems.

2 STRUCTURAL CONNECTIVITY

In the study of the brain, structural connectivity refers to the anatomical connections between brain regions. They are composed mainly of myelinated axons, allowing for fast information transmission. Depending on the level of observation, we can study single-neuron connections (microscale), cortical column connections (mesoscale), or the fiber tracts connecting brain regions (macroscale) (Sporns et al., 2005; Sporns, 2011; Kennedy et al., 2016). We will focus in the macroscale studies as they align more closely to the artificial domain.

The set of structural connections forms the connectome, which is a mathematical model of the physical links between brain regions. The brain is thus considered a network in which each of the regions becomes a node, and the edges are fiber tracts that connect them. This network has a non-random architecture in terms of a scale-free topology (Chung et al., 2017), high modularity (Sporns and Betzel, 2016) and efficiency (Latora and Marchiori, 2001), small worldness (Sporns and Zwi, 2004), and a rich club organization (Van Den Heuvel and Sporns, 2011). These network properties are explained in economic terms because they offer a trade-off between metabolic cost and integration of information (Bullmore and Sporns, 2012; Bassett and Bullmore, 2017).

Structural connections only refer to the possibility of communication between brain regions but say nothing about how much they communicate. It can be said that structure enables function and determines the possible interactions occurring in a system. Thus, the presence of a structural connection enables access to the raw information provided by one processing unit.

From a computational perspective, the structural connectivity in a network system indicates which components sends the information to which others. This concept differs from notion of causality as formally defined in mathematical statistics (Pearl, 2009), which targets cause-effect relations between components using counterfactual interventions. In general is is not possible to perform interventions, so causal relations are notably difficult to infer in a data driven way and requires specific techniques. One example of such techniques is the measure of causal flow that can be constructed using information theory by modifying the mutual information statistic (Ay and Polani, 2008). Yet, a cause-effect relation might not occur between two components even if they exchange information, e.g. such information could be ignored. In AI systems, however, structural connectivity is usually defined and available by design. For example, structural links can be fixed and defined during implementation, or component interactions can be regulated only by a predefined policy. This is the biggest advantage over biological systems like the brain, where the mapping of the entire connectome is an open challenge (Sporns, 2013).

3 FUNCTIONAL CONNECTIVITY

If structural connectivity refers to the physical connections of different nodes, functional connectivity reflects the statistical dependence between their outputs. In the brain, it reflects the level of synchronization over remote regions and have a relationship with co-activation during specific behavior (Biswal et al., 1995; Friston, 2011). Functional connectivity describes how much information is shared by two regions (integration) and, together with the directionality of the information flow, the level of factual access (contrasting to the possible access which is enabled by structural connectivity).

Mimicking the models on structural connectivity, a connectome can be created using functional connectivity to study the functional network's topology. The functional network also exhibits a complex architecture characterised mainly by a power law distribution of connections, high efficiency, a high clustering coefficient and, thus, small-worldness (White et al., 1986; Watts and Strogatz, 1998; Salvador et al., 2005; Valencia et al., 2008). This architecture enables the segregation of structures which become specialized for specific information processing while ensuring their integration in the whole system. The brain can then be divided into functional networks, which are defined as group of highly connected regions which are related to specific cognitive processes (Yeo et al., 2011; Glasser et al., 2016).

Interestingly, there are differences between the structural and functional networks, highlighting the difference between possible access and factual access. Multiple brain models have attempted to study the relation between the two, but the more successful ones are functional connectivity models that are constrained by anatomical connections, such as the Hopf brain model (Deco et al., 2021). The neuroimaging results reveal that a functional organization of statistical relationships can arise from the structural connections within the system. Importantly, this functional organization serves as a more effective indicator of how extensively each component accesses information from other components.

In AI systems, functional connections are traced by the signals, flowing from the sender to the receiving components, whose information is actually integrated by the receivers. The extend of "useful" information can measured by comparing the distributions of components' input and output using information theoretic statistics such as the Conditional Mutual Information, shortened in CMI (Wyner, 1978). This statistic enables to account for the information the receiver component (X) get from the sender (Y) that is redundant with the signal already coming from another source (Z) (Ay and Polani, 2008). Formally, CMI is defined as

$$I(X;Y|Z) = \\ = H(X,Z) + H(Y,Z) - H(Z) - H(X,Y,Z) \quad (1)$$

where $H(X)$ is the entropy (usually Shannon's) of the random variable X (Shannon, 1948).

As the entropy is well defined for any distribution, this measure is applicable without imposing particular restriction to the variables X, Y, Z, and therefore on the system design. In general, CMI can be formulated for variables that can be discrete (ordinal or categorical) or continuous, as long as their density can be suitably estimated. It is applicable to time series (Schreiber, 2000, see Transfer Entropy) and to multivariate variables, although the estimation of their density becomes harder as the dimensions increase, due to the curse of dimensionality (Runge et al., 2012).

3.1 Simulations

To better understand what functional connectivity is, in this section we simulate an AI system where the structural connectivity is fixed and the strength of functional connectivity (the CMI) can be controlled. The simulation was developed to show the benefits of AI systems that are aware of moral values (Steels, 2024; Montes et al., 2022; Abbo and Belpaeme, 2023; Roselli et al., 2023), as part of a research project on information dynamics in social media (Gravino et al., 2022) and moral values (Brugnoli et al., 2024; Marcos-Vidal et al., 2024). In this work, we consider a simple recommender algorithm (Gravino et al., 2019; Marzo et al., 2023) that evaluates if a set of items (e.g., posts from Twitter/X) are fit for given users.

This system consists of four components: a text preprocessing component, a topic detection model (Lavrenko et al., 2002), a moral-topic similarity scorer, and a detector of value dyads from the Moral Foundations Theory (Graham et al., 2013): *Authority/Subversion*, *Care/Harm*, *Fairness/Cheating*, *Loyalty/Betrayal*, *Purity/Degradation* and *No values*. Upon receiving a tweet, the system preprocesses and passes it to the topic detection module, which returns a vector $i_t \in \mathbb{R}^p$ such that i_{tj} is the probability that the tweet t is about topic j and $\sum_j i_{tj} = 1$; the preprocessed text is also passed to values-detector, which returns a vector $i_m \in \mathbb{R}^q$ such that i_{mj} is the probability that the tweet m mainly expresses the moral dyad j and $\sum_j i_{mj} = 1$. Then i_t and i_m are passed to the similarity scorer that calculates a score $r \in [0,1]$ following the formula

$$\lambda S_C(\mathcal{D}(i_m, h_m), u_m) + (1-\lambda) S_C(\mathcal{D}(i_t, h_t), u_t)\,, \quad (2)$$

where:

- S_C is the cosine similarity, to measure the affinity, between the user and the item, on morals and topics;
- $\mathcal{D}(v,h)$, with $v \in \mathbb{R}^k$ for any $k > 0$, is the sample from a Dirichlet distribution with parameters $\alpha_i = 1 + v_i k/(h + 10^{-10})$ for $i = 1, \ldots, k$, enabling the introduction of noise to the input components that is modulated in magnitude by $h \geq 0$;
- The values $h_t \geq 0$ and $h_m \geq 0$ modulate the magnitude of the noise that artificially perturbs the input variables i_t and i_m, respectively;
- $\lambda \in [0,1]$ interpolates between topic (small λ) and moral (large λ) similarities.

In summary, a tweet enters the system which outputs a recommendation score based on the user-post similarities of moral and topics vectors. The parameters λ, u_m, u_t, h_m, h_t must be initialised upon deployment. The architecture is sketched in Figure 1.

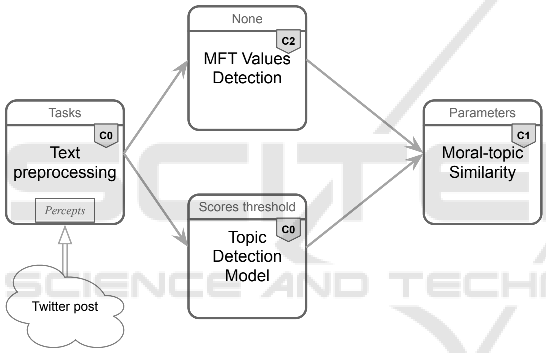

Figure 1: The architecture for the analysis of the CMI.

In this architecture, the CMI lets us quantify how much information the similarity scorer integrates from the value detector, discounting the information already provided by the topic detector. Thus, the CMI is calculated from the outputs of these three components.For notation, we define that from the j-th tweet ($0 \leq j \leq J$), the three components respectively produce the observations of i_t^j, i_m^j and r^j of the random variables Z, X, Y. Since J can be very large (171,067 tweets in the present analysis), we set a batch size of 1000 for the number of observations on which the CMI is calculated, for a total of $B = 171$ batches. So, for each batch b, to obtain the CMI, it suffices to calculate the entropies in (2) from the variables' densities that are estimated via Kernel Density Estimation (KDE) with Gaussian kernel and bandwidth from Scott's method (Scott, 2015). Of note, if the distributions were discrete, one could replace the estimation of the density function with the estimator of the probability mass function by its empirical counterpart.

To understand what the CMI is sensitive to, we set the parameters $h_t = 0$ and $u_t = (1,2,2,2,1,1,2,2,1,1,2,1,1,2,1,1,2,2,1,1,2,2,2,2,2,1)/40$ by assigning weight 1 or 2 uniformly at random to every topic and then renormalising. The empirical distributions of the CMI are thus generated on B batches from the architecture simulations with:

- $\lambda \in \{0, 0.1, 0.2, 0.5, 0.7, 1\}$, to weight the input from the moral component;
- $h_m \in \{0, 0.01, 1, 10, 10^{10}\}$, to perturb the input from the moral component;
- $u_m \in \{(1,1,1,1,1,1)/6, (1,2,2,1,1,2)/9, (1,0,0,0,0,0)\}$, to simulate a user with uniform, balanced and extreme moral preferences;

The analysis of Figure 2 shows that the CMI between the value detector and the similarity scorer, conditioned on the topic detector, decreases with λ and becomes null when $\lambda = 0$. This transition reflects the reduction of new information, from the moral component, integrated in the moral-topic similarity. The plot also show that the information transfer gets corrupted as the output from the moral component becomes noisier with increasing h_m.These expected trends hold true irrespective of the user's moral preferences u_m, with the case $\lambda = 1$ dominating the others for every fixed noise level. Interestingly, when the noise is low, the CMI distribution is quite different between users. CMI increases from uniform to balanced and then to extreme profiles, suggesting that the similarity scorer adapts its behaviour to the users'. This evidence could be the result of the distribution of Y becoming less variable and, therefore, less entropic as the user's moral profile becomes more extreme.

In practice, with this measure, we can rank different systems to select the architecture specifications with stronger functional connectivity for selected components. Another application of the measure is to support design decisions in the presence of tradeoffs between performances and functional connectivity. For example, the similarity scorer component always provides the highest integration of the moral detection component when $\lambda = 1$, at the cost of totally ignoring the input from the topic detection model, which is quite unreasonable. However, comparing the CMI across different values of λ, it becomes apparent that values of $\lambda = 0.7$ or $\lambda = 0.5$ do not degrade considerably the amount of integrated information, indicating these as acceptable design choices.

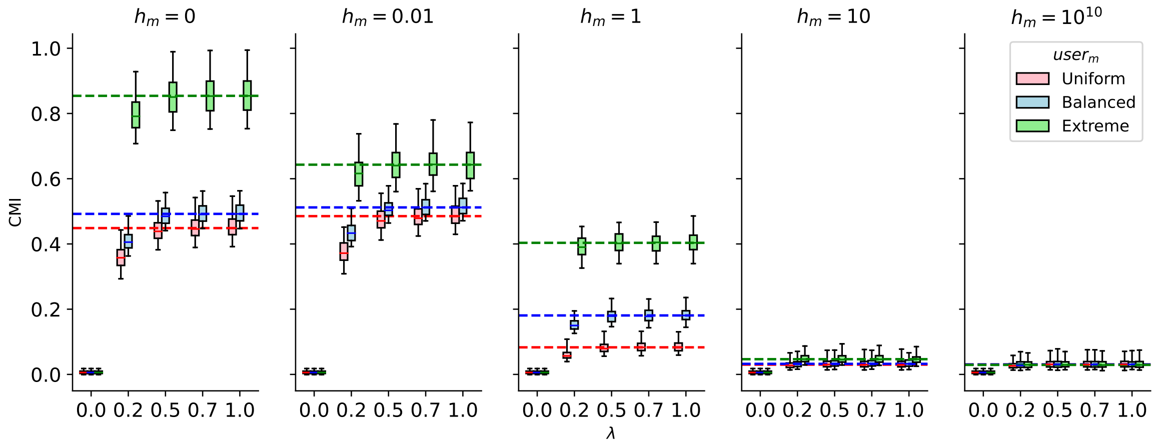

Figure 2: Comparison of integration measure from the simulations of different architectures. Each boxplot summarises the distribution of CMI between the value detector and the similarity scorer conditioned on the topic detector from 171 batches of observations, with boxes covering the interquartile range and containing a full line for the median and whiskers at the 5th and 95th percentiles. Boxes are grouped in triplets, colour-coded by $user_m$ choice as in the legend, and horizontally located depending on the parameters λ (within panels) and h_m (between panels) chosen for their simulation. Dashed horizontal lines are placed at the level of the medians of each user's boxplot with $\lambda = 1$ and coloured as the related box.

4 OPERATIONAL CONNECTIVITY

As a system processes an input, a series of components get activated, generating a path in the network. The ensemble of these paths is the target of the operational connectivity analysis, an approach commonly employed to study the brain (Casali et al., 2013). The brain's activity can be sampled using techniques such as functional magnetic resonance imaging or electroencephalogram, and data from these trials are often used to infer the structural connectivity and the functional connectivity of the brain (Bullmore and Sporns, 2009).

By investigating the chain reactions of neuronal activity, also called neuronal avalanches, a study found that their size and lifetime follow a power law distribution, whose scale-free property is a trait of scalable, self-organising systems (Beggs and Plenz, 2003). The same authors proposed that the power law structure emerges from the avalanches behaving like a branching process (Harris et al., 1963). This is a probabilistic process to model the reproduction of, for example, the activity of a neuron that triggers a stochastic number of neighbouring neurons. In particular, the branching coefficient, optimal for the observed dynamic, showed that the process is at a critical state, meaning that the amount of information transmitted by an avalanche is maintained high while preventing the catastrophic activation of the entire system (occurring at the supercritical state).

High scalability and stable information transmission are desirable properties for AI systems that evolve over time. The analysis of operational connectivity, as in (Beggs and Plenz, 2003), provides ways to assess where and how single datapoints percolate through the system. In particular, the coefficients from branching process theory and power law distributions enable us to measure how scalable the data transmission process is. However, this analysis is only sensible for those AI systems whose components may turn on and off over time with no predictable or preprogrammed schedule. Trivially, a convolutional feed-forward deep neural network activates all its layers of components irrespective of the input.

As different inputs can be processed similarly by the system, their sequences of activation might share the same common patterns. For example, an input generating a long path that touches several regions of the system is indicative of an expensive process, which becomes outstanding if the path observed is particularly unusual compared to the previous others. Also, a path that gets stuck into an infinite loop bouncing between the same set of components might flag an internal clash that calls for a mechanism to solve the conflict.

We contend that investigating the operational connectivity by measuring avalanches and patterns of activity can contribute to understanding both the system and the salience of its inputs.

5 CONCLUSIONS

We have analysed access of information, a hallmark of awareness, from both neurological and computational perspectives. For systems whose components form a network structure, three distinct types of connectivity were identified: structural, defining where information can flow; functional, defining where novel information flows; operational, traced by the series of components activation triggered by a single input. Finally, we reviewed some measures for access of information defined from the information flow on the different types of connectivity. In particular we focused on the quantification of functional connectivity via Conditional Mutual Information, and showed with simulations the utility of this metric to evaluate access of information in AI systems.

By focusing on information access, we did not examine other aspects of awareness. When information is internalised, for example, the content of information is also elaborated, integrating the processes of several components. Providing rigorous quantification for all these aspects might benefit the development of aware AI systems. Meanwhile, the synthesis of neurological and computational perspectives will contribute to a broader understanding of awareness in both biological and artificial systems.

ACKNOWLEDGEMENTS

This work has been supported by the Horizon Europe VALAWAI project (grant agreement number 101070930). We thank Prof. Luc Steels for the fruitful discussions that were essential to conceiving the manuscript.

REFERENCES

Abbo, G. A. and Belpaeme, T. (2023). Users' perspectives on value awareness in social robots. In *HRI2023, the 18th ACM/IEEE International Conference on Human-Robot Interaction*, pages 1–5.

Ay, N. and Polani, D. (2008). Information flows in causal networks. *Advances in complex systems*, 11(01):17–41.

Bassett, D. S. and Bullmore, E. T. (2017). Small-world brain networks revisited. *The Neuroscientist*, 23(5):499–516.

Beggs, J. M. and Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *Journal of neuroscience*, 23(35):11167–11177.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and brain sciences*, 18(2):227–247.

Brugnoli, E., Gravino, P., and Prevedello, G. (to appear in 2024). Moral values in social media for disinformation and hate speech analysis. In *VALE workshop Proceedings*, LNAI.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186–198.

Bullmore, E. and Sporns, O. (2012). The economy of brain network organization. *Nature reviews neuroscience*, 13(5):336–349.

Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Casarotto, S., Bruno, M.-A., Laureys, S., Tononi, G., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Science translational medicine*, 5(198):198ra105–198ra105.

Chalmers, D. J. (1997). *The conscious mind: In search of a fundamental theory*. Oxford Paperbacks.

Chung, M. K., Hanson, J. L., Adluru, N., Alexander, A. L., Davidson, R. J., and Pollak, S. D. (2017). Integrative structural brain network analysis in diffusion tensor imaging. *Brain Connectivity*, 7(6):331–346.

Deco, G., Vidaurre, D., and Kringelbach, M. L. (2021). Revisiting the global workspace orchestrating the hierarchical organization of the human brain. *Nature human behaviour*, 5(4):497–511.

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain connectivity*, 1(1):13–36.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., and Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology*, volume 47, pages 55–130. Elsevier.

Gravino, P., Monechi, B., and Loreto, V. (2019). Towards novelty-driven recommender systems. *Comptes Rendus Physique*, 20(4):371–379.

Gravino, P., Prevedello, G., Galletti, M., and Loreto, V. (2022). The supply and demand of news during covid-19 and assessment of questionable sources production. *Nature human behaviour*, 6(8):1069–1078.

Harris, T. E. et al. (1963). *The theory of branching processes*, volume 6. Springer Berlin.

Kennedy, H., Van Essen, D. C., and Christen, Y. (2016). *Micro-, meso-and macro-connectomics of the brain*. Springer Nature.

Koster, R., Balaguer, J., Tacchetti, A., Weinstein, A., Zhu, T., Hauser, O., Williams, D., Campbell-Gillingham, L., Thacker, P., Botvinick, M., et al. (2022). Human-centred mechanism design with democratic ai. *Nature Human Behaviour*, 6(10):1398–1407.

Latora, V. and Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical review letters*, 87(19):198701.

Lavrenko, V., Allan, J., DeGuzman, E., LaFlamme, D., Pollard, V., and Thomas, S. (2002). Relevance models for topic detection and tracking. In *Proceedings of the human language technology conference (HLT)*, pages 104–110.

Marcos-Vidal, L., Marchesi, S., Wykowska, A., and Pretus, C. (to appear in 2024). Moral agents as relational systems: The contract-based model of moral cognition for ai. In *VALE workshop Proceedings*, LNAI.

Marzo, G. D., Gravino, P., and Loreto, V. (2023). Recommender systems may enhance the discovery of novelties.

Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., Agirre, E., Heintz, I., and Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys*, 56(2):1–40.

Montes, N., Osman, N., and Sierra, C. (2022). A computational model of ostrom's institutional analysis and development framework. *Artificial Intelligence*, 311:103756.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, 3(none):96 – 146.

Roselli, C., Marchesi, S., De Tommaso, D., and Wykowska, A. (2023). The role of prior exposure in the likelihood of adopting the intentional stance toward a humanoid robot. *Paladyn, Journal of Behavioral Robotics*, 14(1):20220103.

Runge, J., Heitzig, J., Petoukhov, V., and Kurths, J. (2012). Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Physical review letters*, 108(25):258701.

Salvador, R., Suckling, J., Coleman, M. R., Pickard, J. D., Menon, D., and Bullmore, E. (2005). Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral cortex*, 15(9):1332–1342.

Schreiber, T. (2000). Measuring information transfer. *Physical review letters*, 85(2):461.

Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

Sporns, O. (2011). The human connectome: a complex network. *Annals of the new York Academy of Sciences*, 1224(1):109–125.

Sporns, O. (2013). The human connectome: origins and challenges. *Neuroimage*, 80:53–61.

Sporns, O. and Betzel, R. F. (2016). Modular brain networks. *Annual review of psychology*, 67:613–640.

Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS computational biology*, 1(4):e42.

Sporns, O. and Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics*, 2:145–162.

Steels, L. (to appear in 2024). Values, norms and ai. introduction to the vale workshop. In *VALE workshop Proceedings*, LNAI.

Valencia, M., Martinerie, J., Dupont, S., and Chavez, M. (2008). Dynamic small-world behavior in functional brain networks unveiled by an event-related networks approach. *Physical Review E*, 77(5):050905.

Van Den Heuvel, M. P. and Sporns, O. (2011). Rich-club organization of the human connectome. *Journal of Neuroscience*, 31(44):15775–15786.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442.

White, J. G., Southgate, E., Thomson, J. N., Brenner, S., et al. (1986). The structure of the nervous system of the nematode caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci*, 314(1165):1–340.

Wyner, A. D. (1978). A definition of conditional mutual information for arbitrary ensembles. *Information and Control*, 38(1):51–59.

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*.