Charlie Gross:

an inspiration

David Perrett

Date of deposit	22 10 2020
Document version	Author's accepted manuscript
Access rights	
Citation for published version	Perrett, D 2020, 'Charlie Gross: an inspiration', *Progress in Neurobiology*, vol. in press, 101928.
Link to published version	https://doi.org/10.1016/j.pneurobio.2020.101928

Charlie Gross: An Inspiration

I want to express my profound thanks to Charlie Gross. His findings were truly inspirational for me as a student and a researcher. My purpose in writing this is to acknowledge the influence Charlie played in shaping my own work and, I think, in making the brain an understandable organ.

It was in 1973 as an undergraduate in a neuropsychology class that I first heard about the brain cell that Charlie and co-workers (Gross et al. 1972) had discovered responding to the sight of a hand. Such findings were astonishing. Neuropsychology had long suggested that specific types of recognition were conducted in specific brain systems but it was unknown how recognition was achieved. Charlie's discoveries implied that the brain was computing the form of objects and that this process was realized (explicitly) in the activity of its working components, the individual neurones. In short, recognition could be understood if it was possible to define what individual cells were responding to along the visual processing pathway.

3 years later (in 1977) as a postgraduate student in Berkeley, California, it was evident to methat more than a few prominent visual scientists did not believe Charlie's conclusions. Instead, their interpretation of what Charlie had described in the temporal cortex was a cell response to a vertically oriented spatially repetitive pattern of 0.5 cycles per degree. After all, cells tuned to different orientations and spatial frequencies were well described in the early visual cortex (e.g. De Valois et al. 1978). Spatial frequency tuning rather than tuning for a hand implied little change in the processing across early to late stages of visual processing. Was the brain really that wasteful of expensive energy computing the same thing over and over again?

Models of recognition

In the 1970s there was a belief that the early visual cortex performed a Fourier decomposition into component spatial frequencies at different orientations. What happened next in terms of visual processing was mysterious and remained so for some time after Charlie's publication. Some suggested that the component frequency information was used in the holographic reconstruction (e.g. Pribram & Meade 1999; Pribram 1991) of objects within the brain. A holographic record of an object does allow the formation of a 3D image of the object but reconstituting an object's image merely postpones the job of recognition. Taking a 2D photo of a beast with two horns does not tell you that you are looking at an elk or a cow or even that it is an animal. Personally, I found this holographic account impossible to fathom; it did not seem to be a model of vision so much as an analogy. A practical theory of visual recognition needs to specify how to make explicit information about object edges and components (Marr 1982).

Charlie's discoveries meant that brain processing was tangible and understandable. Decades later we can see his findings as supporting a hierarchical increase in the complexity of visual stimuli that cells respond to coupled with an increase in the generalization or invariance in responses to the same stimuli across space. The hierarchy started with simple features, progressing to more elaborate features (Tanaka 1996) and culminating in very complex patterns such as hands and faces within the temporal cortex (Oram and Perrett 1994). Such a hierarchy of visual feature analysis and invariance is evident in recent artificial neural networks trained to categorize objects from images (Cadieu et al. 2014; Khaligh-Razavi et al. 2014; Yamins et al. 2014; Güçlü & van Gerven 2015; Cichy et al. 2016).

Replicating Charlie's findings and Lady Luck

In the late 1970's the data on selectivity for complex stimuli from Charlie's lab was sparse: one cell was described responding to a hand and there was also mention of similar cell selectivity for faces but no detailed description. Charlie's descriptions intrigued me but they needed replication and such cells were worthy of systematic study. I suggested to George Ettlinger in London that such cells would be a suitable subject to study for a PhD (Riddley et al. 1977). In the end, I had an opportunity to pursue a PhD with Edmund Rolls in Oxford.

An intriguing part of Charlie's observations was their serendipity. He notes "having failed to drive a unit with any light stimulus, we waved a hand at the stimulus screen and elicited a very vigorous response from the previously unresponsive neuron" (Gross 2008). Such chance encounters might suggest that my drive to study such cells was doomed to failure because of their rarity and selectivity. In the first years of my doctoral studies we recorded in many brain areas (parietal cortex, caudate nucleus and hippocampus) but I had not forgotten thetemporal cortex. As luck would have it, I was fated (similarly to Charlie) and to my great surprise while working with an unresponsive cell in the tail of caudate nucleus I found that it responded to the sight of my face. It turned out that I was off-target and recording in the superior temporal sulcus within a cortical region containing many cells responsive to faces (Bruce et al. 1981; Harries and Perrett 1991; Tsao et al. 2006). It was a dream come true to be able to return to the same region of cortex and make a systematic study of such cells (Perrett et al. 1982). I realized then that I would spend the rest of my career working with faces as there were so many questions to address.

I met Charlie at a handful of conferences. Each time he proved as unconventional as his work. In Israel at Masada virtually all visitors chose a cable-car to reach the hill-top fortress whereas Charlie chose the Snake path and an hour's climb in the midday heat. His maverick approach in all was good to follow; it allowed for heretical results! In Florida where I described cells selective for face identity, Charlie eschewed public transport, taxis and rental cars and chose to hitch rides - an action common for students but very rare for faculty.

Generalizing recognition across views

Charlie's work revealed several cell properties that I think help to de-mystify the process of neural recognition. The cell responsive to a hand that he described was orientation sensitive (responding to the sight of fingers pointing upwards but not to fingers pointing downwards). Orientation specificity was found to be common amongst cells in this part of the brain. Here was the first clue as to how the brain was processing objects. Tom Collett from Sussex pointed out to me that, rather than computing one 3D description of an object as suggested by Marr (1982), the brain was building descriptions that were specific to different vantage points. It turned out that most cells responsive to faces were also sensitive to orientation, to image size and to perspective view (Desimone et al. 1984; Perrett and Oram 1998).

Working first with perspective view we realized that generalization across perspective view (front, side or back of the head) can be achieved by pooling the outputs of cells sensitive to different views of the same object (Perrett et al. 1984, 1985). Likewise, generalization across the different image sizes of an object seen at different distances can be achieved by pooling the outputs of cells tuned to the same object but selective for image size (Perrett and Oram 1998). By the same pooling principle, recognition independent of orientation can depend on pooling across outputs of cells tuned to different orientations relative to gravity. Moreover, generalization across different parts of the same object can arise simply from combining the outputs of cells sensitive to specific parts (Wachsmuth et al. 1994).

Rather than computing one all-encompassing object description, the brain performs a more straightforward and limited set of operations each of which specifies the appearance of a meaningful object from a limited range of views. The descriptions of different body parts (face hand, torso) from a given orientation and perspective view are not only useful for recognizing an object as a body but they also allow specification of what the body is doing (Perrett et al. 1984, 1989). Specific postures imply specific actions (Jellema and Perrett 2003a, 2003b) and an analysis of the disposition of body parts allows the intentions of others to be realized (Jellema et al. 2000).

Another property Charlie and co-workers described in the temporal cortex was the large receptive fields of cells which usually included the fovea and extended over a quadrant or more of the visual field. Both his lab and our own emphasized the extent of the receptive fields which allows generalization of recognition across retinal position. Yet since the receptive fields do not extend over the total visual world, some positional information remains available in responses; probably more than we originally realized. Such information in the temporal cortex and its inputs may allow specification of the position of a hand relative to 'goal' objects (Fleischer et al. 2013), underlying the selectivity shown by neurones responsive to the sight of particular hand actions in the temporal cortex (Perrett et al. 1989; 1990) and 'mirror neurones' in premotor and parietal cortex (Di Pellegrino et al. 1992; Fogassi et al. 2005). Indeed, the temporal cortex responses to hand actions (Perrett et al. 1990) led me to pester Giacomo Rizolatti as to whether premotor cells active during grasping also responded to the sight of hands grasping. In this indirect way Charlie's work contributed to the description of mirror neurone properties.

Conclusion

Charlie's work, particularly his description of a cell responsive to the sight of a hand, were pivotal for me. They inspired my physiological studies and entire career working on recognition. The wider significance of his work was in kick-starting a cascade of physiological findings that provided the milieu for developing biologically based computational models of recognition (Riesenhuber and Poggio 1999; Ullman et al. 2002; Serre et al. 2007; Cadieu et al. 2014; Khaligh-Razavi et al. 2014; Yamins et al. 2014; Güçlü & van Gerven 2015; Cichy et al. 2016). The success of artificial recognition stemming from his work has repercussions that are transforming society. One cell made all the difference.

References

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, 46(2), 369-384. DOI: https://doi.org/10.1152/jn.1981.46.2.369

Cadieu, C. F., Hong, H., Yamins, D. L., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N.J. & DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Computational Biology*, 10(12), e1003963. DOI: https://doi.org/10.1371/journal.pcbi.1003963

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, 6, 27755. DOI: https://doi.org/10.1038/srep27755

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. Journal of Neuroscience, 4(8), 2051-2062. DOI: https://doi.org/10.1523/JNEUROSCI.04-08-02051.1984

De Valois, R.L., Albrecht, D.G., Thorell, L.G. (1978). Cortical cells: bar and edge detectors, or spatial frequency filters? In: Cool S.J., Smith E.L. (eds) Frontiers in Visual Science. Springer Series in Optical Sciences, vol 8. Springer, Berlin, Heidelberg. DOI https://doi.org/10.1007/978-3-540-35397-3_49

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91(1), 176-180. DOI: https://doi.org/10.1007/bf00230027

Fleischer, F., Caggiano, V., Thier, P., & Giese, M. A. (2013). Physiologically inspired model for the visual recognition of transitive hand actions. *Journal of Neuroscience*, 33(15), 6563-6580. DOI: https://doi.org/10.1523/JNEUROSCI.4129-12.2013

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science*, 308(5722), 662-667. DOI: https://doi.org/10.1126/science.1106138

Güçlü, U., & van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, 35(27), 10005-10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Gross, C. G. (2008). Single neuron studies of inferior temporal cortex. *Neuropsychologia*, 46(3), 841-852. DOI: 10.1016/j.neuropsychologia.2007.11.009

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. Journal of Neurophysiology, 35(1), 96-111. DOI: https://doi.org/10.1152/jn.1972.35.1.96

Harries, M. H., & Perrett, D. I. (1991). Visual processing of faces in temporal cortex: Physiological evidence for a modular organization and possible anatomical correlates. *Journal of Cognitive Neuroscience*, 3(1), 9-24. DOI: https://doi.org/10.1162/jocn.1991.3.1.9.

Jellema, T., Baker, C. I., Wicker, B., & Perrett, D. I. (2000). Neural representation for the perception of the intentionality of actions. Brain and Cognition, 44(2), 280-302. DOI: https://doi.org/10.1006/brcg.2000.1231

Jellema, T., & Perrett, D. I. (2003a). Perceptual history influences neural responses to face and body postures. *Journal of Cognitive Neuroscience*, 15(7), 961-971. DOI: https://doi.org/10.1162/089892903770007353

Jellema, T., & Perrett, D. I. (2003b). Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion? *Neuropsychologia*, 41(13), 1728-1737. https://doi.org/10.1016/S0028-3932(03)00175-1

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, 10(11), e1003915. https://doi.org/10.1371/journal.pcbi.1003915

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Publisher WH Freeman Company.

Oram, M. W., & Perrett, D. I. (1994). Modeling visual recognition from neurobiological constraints. *Neural Networks*, 7(6-7), 945-972. https://doi.org/10.1016/S0893-6080(05)80153-4

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, 47(3), 329-342. DOI: https://doi.org/10.1007/bf00239352

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D. & Jeeves, M. A. (1984). Neurons responsive to faces in the temporal cortex: Studies of functional organization, sensitivity to identity and relation to perception. *Human Neurobiology*, 3, 197-208. PMID: 6526706

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D. & Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London* B, 223, 293-317. DOI: https://doi.org/10.1098/rspb.1985.0003

Perrett, D. I., Harries, M. H., Bevan, R., Thomas, S., Benson, P. J., Mistlin, A. J., Chitty, A. J., Hietanen, J. K., & Ortega, J. E. (1989). Frameworks of analysis for the neural representation of animate objects and actions. *Journal of Experimental Biology*, 146(1), 87-113. PMID: 2689570

Perrett, D.I., Mistlin, A.J., Harries, M.H. and Chitty, A.J. (1990). Understanding the visual appearance and consequences of hand actions. In: Vision and Action: The Control of Grasping (Ed. M.A. Goodale). Norwood, N.J.: Ablex Publishing Corp. pp 163-180.

Perrett, D. I., & Oram, M. W. (1998). Visual recognition based on temporal cortex cells: Viewer-centred processing of pattern configuration. *Zeitschrift für Naturforschung* C, 53(7-8), 518-541. DOI: https://doi.org/10.1515/znc-1998-7-807

Pribram K.H. (1991). *Brain and Perception: Holonomy and Structure in Figural Processing. eBook Published* 2013 New York. https://doi.org/10.4324/9780203728390.

Pribram K. H., & Meade S. D. (1999). Conscious awareness: Processing in the synaptodendritic web. *New Ideas in Psychology,* 17 (3): 205–214. DOI: https://doi.org/10.1016/S0732-118X(99)00024-0

Ridley, R. M., Hester, N. S., & Ettlinger, G. (1977). Stimulus-and response-dependent units from the occipital and temporal lobes of the unanaesthetized monkey performing learnt visual tasks. *Experimental Brain Research*, 27(5), 539-552. DOI: https://doi.org/10.1007/bf00239042

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11), 1019 - 1025. DOI: https://doi.org/10.1038/14819

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (3), 411-426. https://doi.org/10.1109/tpami.2007.56

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, 19(1), 109-139. DOI: https://doi.org/10.1146/annurev.ne.19.030196.000545

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, 311(5761), 670-674. DOI: https://doi.org/10.1126/science.1119983

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5(7), 682-687. DOI: https://doi.org/10.1038/nn870

Wachsmuth, E., Oram, M.W., & Perrett, D.I. (1994). Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. *Cerebral Cortex*, 4, 509-522. DOI: https://doi.org/10.1093/cercor/4.5.509

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111(23), 8619-8624. https://doi.org/10.1073/pnas.1403112111