COMMENTARY

Searching for the neural causes of criminal behavior

Ralph Adolphs[a,b,1], Jan Gläscher[c], and Daniel Tranel[d,e]

All behavior is proximally caused by the brain, but the neural causes of most complex behaviors are still not understood. Much of our ignorance stems from the fact that complex behavior depends on distributed neural control. Unlike a reflex, where the arc from sensation to action can be traced through a few synapses, most volitional behavior involves a dense causal web through which stimuli, memories, beliefs, and other factors exert their effects. Disruption anywhere in this causal web can produce effects that are difficult to trace back to their origin. Against this background, the finding that focal lesions of the ventromedial prefrontal cortex could lead to immoral and even criminal behavior generated considerable surprise and interest (1, 2). While a number of rare cases have now been described in whom a focal lesion caused criminality, these are neither very consistent (the lesions occur in several different anatomical locations) nor at all reliable (only a small fraction of patients, for any lesion location, show criminal behavior). To explain the effects of a lesion on criminal behavior, we need to understand what it is that the lesion does to the rest of the brain, a network-level understanding of lesion effects now provided by the new study of Darby et al. in PNAS (3).

A Disconnection Approach

Darby et al. (3) began by searching the literature for cases documenting criminal behavior following a focal brain lesion. They found 17 of these, with lesions predominantly including the prefrontal cortex but also other regions. To estimate the distal effects of such brain damage, lesion network mapping (4) was used, showing which other brain regions, in healthy brains, would normally be functionally connected with the location of the lesion. This approach effectively treats the brain of the lesion patient as equivalent to a healthy brain in which all those regions functionally connected to the lesion site have been instantaneously affected. To do this, Darby et al. (3) leveraged a large reference dataset of resting-state functional MRI (fMRI) data from healthy individuals [from the Brain Genomics Superstruct Project (5)]. From each lesion patient a map was produced showing where, in this resting-state dataset of healthy individuals, the lesion site had strong functional connectivity. Maps of these functionally connected regions were then overlapped across the patients, producing a composite map showing all of the places in a healthy brain that were predicted to be impacted by the lesions. This innovative approach (4, 6) takes seriously the notion of a disconnection syndrome, a term popularized by Norman Geschwind (7) in the 1960s: it is the idea that behavioral impairments can be caused by the disruption of the connections between brain regions rather than by damage to one particular region itself.

The composite connectivity map based on positive correlations in the resting-state dataset in fact included many of the regions that, when themselves lesioned, were associated with criminality: the ventromedial and orbital prefrontal cortex, as well as the anterior temporal lobes. In contrast, the dorsolateral prefrontal cortex was negatively functionally connected with the lesions. This pattern of functional connectivity was not seen for 63 other lesions (none of which had any effect on criminality). But it overlapped with brain regions found in task fMRI studies to be activated when healthy participants were confronted with personal moral dilemmas involving harm to another person. Finally, Darby et al. (3) took care to replicate their findings in an independent sample of 23 patients whose lesions were also implicated in criminal behavior, but where the temporal association between the lesion and changes in behavior was uncertain. Taken together, and in the context of prior lesion (2) and activation studies (8–10) of moral and criminal behavior, the results (3) offer a very intriguing addition not only to the list of brain regions in which damage can alter behavior related to criminality, but also to the mechanisms by which they do so.

Reorganization and Individual Differences

However, there are several serious limitations to the work, as Darby et al. (3) are careful to discuss in their

[a]Division of Humanities and Social Science, California Institute of Technology, Pasadena, CA 91125; [b]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125; [c]Institute for Systems Neuroscience, University Medical Center Hamburg–Eppendorf, 20246 Hamburg, Germany; [d]Department of Neurology, University of Iowa College of Medicine, Iowa City, IA 52242; and [e]Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA 52242

Author contributions: R.A., J.G., and D.T. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 601.

[1]To whom correspondence should be addressed. Email: radolphs@caltech.edu.

paper. First and foremost among these is the lack of any functional imaging in the lesion patients themselves. All of the inferences about distal functional effects of the lesion are derived from assumptions based on functional connectivity in a database of healthy brains. However, it is well known that functional connectivity changes in psychiatric and neurological disorders (11, 12), and it would seem likely that there are substantial changes in the functional connectome of the lesion patients, especially given that reorganization and compensation will no doubt play a role. Even instantaneous changes in brain networks are likely to be considerably more complex than can be revealed with the resting-state connectivity approach taken in the present paper (3). For example, acute pharmacological inactivation of the amygdala in monkeys causes dramatic changes in network architecture that include changes in functional connectivity between distal brain regions (13).

Perhaps most worrisome is the neglect of individual differences. Not only will there be premorbid individual differences in personality and other variables that interact with the effect of a lesion, but (probably in good part due to this) there is also the fact that most patients who have a lesion in one of the regions implicated by Darby et al.'s (3) analyses (as causally linked to criminal behavior) do not exhibit criminal behavior. It would be of the utmost importance to obtain resting-state fMRI data in criminal lesion cases, such as those described by Darby et al., because it is known that individual functional connectomes not only differ from one another, but also from a group average, such as the one used in their study (14). The most valuable comparisons would then be with two other datasets: with the same individual, but before the lesion (notoriously difficult to obtain since the lesions are generally unpredictable accidents of nature); and with other lesion patients who have damage in the same region, but don't show criminal behavior. Understanding the risks, mechanisms, and potential for brain-targeted treatments of criminal behavior will certainly require attention to single individuals, a trend now followed by much of neuroimaging (15).

Moral Responsibility and Free Will

The Darby et al. study (3) offers the valuable hypothesis of a network of core regions that are most closely associated with criminal behavior (the set revealed by the lesion network mapping). But what exactly is criminal behavior, conceived in terms of cognitive processes? Darby et al. probe this issue by examining the overlap of these core criminality-associated regions with brain regions inferred from functional imaging data to subserve processes that might well come into play during criminal behavior. Indeed, overlap was found with brain regions known to be important for theory of mind and value-based decision-making, a finding also consistent with prior studies showing deficits in these abilities in patients whose lesions overlap with the ones in the present study (16). How much this partial decomposition begins to explain criminality, however, seems still very much open to question because, again, the vast majority of patients with impairments in theory of mind or value-based decision-making are not criminals. In a way this is good news for folk psychology and jurisprudence: our concepts of moral responsibility and free will are not challenged by these data, since neither the brain lesions nor the lesion network maps explain criminality (17).

Presumably this is so because the many other ingredients that need to come into play refer to factors outside the brain. Genes, upbringing, provocation, alcohol and drugs, and other factors that cause momentary emotions and lapses in control, are all going to act through the brain but may not be easily mapped onto the brain. Only by gaining a firm handle on these other factors can we understand the substrate on which a focal brain lesion could cause criminal behavior.

This fact is particularly important because laypeople and the media alike continually search for objective explanations of criminal behavior, when in fact criminality is highly relative to particular laws and the interpretation of behaviors in a specific context. In a supplemental table to the Darby et al. study (3), the authors list the examples of criminality that they included, some of which are not obviously so (lying, theft, fraud), and some of which go in the opposite direction (in two patients the lesion caused the cessation of prelesion criminal behavior). A premeditated white-collar crime and a murder committed in blind fury may have few psychological features in common, making it more fruitful to treat them as separate behaviors to try to understand, than as aspects of a single very heterogeneous category we should attempt to investigate. A more precise operationalization of criminality, and a much better understanding of its psychological causes, are likely to be a prerequisite for understanding the neurological causes.

Acknowledgments

The authors were supported by National Institute of Mental Health Grant 2P50MH094258.

1 Damasio H, Grabowski T, Frank R, Galaburda AM, Damasio AR (1994) The return of Phineas Gage: Clues about the brain from the skull of a famous patient. *Science* 264:1102–1105.
2 Anderson SW, Bechara A, Damasio H, Tranel D, Damasio AR (1999) Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nat Neurosci* 2:1032–1037.
3 Darby RR, Horn A, Cushman F, Fox MD (2018) Lesion network localization of criminal behavior. *Proc Natl Acad Sci USA* 115:601–606.
4 Boes AD, et al. (2015) Network localization of neurological symptoms from focal brain lesions. *Brain* 138:3061–3075.
5 Holmes AJ, et al. (2015) Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures. *Sci Data* 2:150031.
6 Sutterer MJ, et al. (2016) Canceled connections: Lesion-derived network mapping helps explain differences in performance on a complex decision-making task. *Cortex* 78:31–43.
7 Geschwind N (1965) Disconnexion syndromes in animals and man. I. *Brain* 88:237–294.
8 Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD (2004) The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44:389–400.
9 Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD (2001) An fMRI investigation of emotional engagement in moral judgment. *Science* 293:2105–2108.
10 Moll J, Zahn R, de Oliveira-Souza R, Krueger F, Grafman J (2005) Opinion: The neural basis of human moral cognition. *Nat Rev Neurosci* 6:799–809.
11 Greicius M (2008) Resting-state functional connectivity in neuropsychiatric disorders. *Curr Opin Neurol* 21:424–430.
12 Greicius MD, Kimmel DL (2012) Neuroimaging insights into network-based neurodegeneration. *Curr Opin Neurol* 25:727–734.
13 Grayson DS, et al. (2016) The Rhesus monkey connectome predicts disrupted functional networks resulting from pharmacogenetic inactivation of the amygdala. *Neuron* 91:453–466.
14 Gordon EM, et al. (2017) Precision functional mapping of individual human brains. *Neuron* 95:791–807.e7.
15 Dubois J, Adolphs R (2016) Building a science of individual differences from fMRI. *Trends Cogn Sci* 20:425–443.
16 Bechara A, Damasio H, Damasio AR (2000) Emotion, decision making and the orbitofrontal cortex. *Cereb Cortex* 10:295–307.
17 Roskies A (2006) Neuroscientific challenges to free will and responsibility. *Trends Cogn Sci* 10:419–423.