

3.2 Comments on Adina Roskies, "Can Neuroscience Resolve Issues about Free Will?"

Michael N. Shadlen

In "Can Neuroscience Resolve Issues about Free Will?" Adina Roskies knits together traditional and modern philosophical ideas with emerging concepts arising from the neuroscience of decision making. It is an elegant and nuanced synthesis. Roskies feels the neurobiology of decision making offers insight into the machine behind the mind while arguing that mechanism does not undermine mindedness or threaten cherished ideas. While she believes that the neuroscience bears on traditional philosophical debates about free will and responsibility, its role is limited. In particular, neuroscience cannot settle fundamental issues like the truth of determinism. To her fellow compatibilists, her message is that understanding the machinery is relevant. Her message to libertarians is that the neurobiology supports indeterminism, although not the right kind of indeterminism. I think she has more to offer libertarians, as I will explain.

Roskies's grasp of the neuroscience is impressive, leaving me very little to say. However, two themes might benefit from elaboration. The first concerns a component of the neural mechanism that I will refer to as a policy: a higher-level setting that governs the basic mechanism for converting evidence to a decision. The second concerns the nature of neural noise. These points are likely to be most interesting to libertarians and hard incompatibilists, but let me preface this with an appeal to Roskies's fellow compatibilists.

The most compelling arguments for compatibilism would render the neurobiology as irrelevant as the proposition of determinism itself. From the perspective of many writers, ethical responsibility does not rest on the validity of determinism or many other philosophical propositions. Instead, ethics lies beyond the limits of propositional construction; ethics rests on a form of contractualism (Williams, 1985) between human beings who choose (or are built) to socialize. Nonetheless, mental capacities are often relevant in considerations of culpability, credit, regret, and the like. At the

very least, the neuroscience of decision making is relevant to an assessment of relevant mental capacities. I think Roskies expresses a similar view.

For libertarians, Roskies offers a way out of the “luck problem.” She exposes a form of indeterminacy that arises from noisiness, but instead of relying on this (weak, in her view) indeterminacy to ground freedom, she mentions other higher-level aspects of the decision-making mechanism that control the process of deliberation toward a choice, what might be termed policies (Shadlen & Roskies, 2012). An example is the trade-off between the speed and accuracy of a decision. The mechanism is essentially a brain setting that establishes a threshold for terminating a decision. The threshold operates on the neural representation of accumulated evidence or elapsed decision time or some combination of the two (Gold & Shadlen, 2007).¹

Policies instantiate value in the decision process. Speed–accuracy is a special case where the value (or cost) of time is pitted against the value of accuracy. Other examples of policies would include the relative weighting of different sources of evidence, the value assigned to potential outcomes, temporal discounting of such valuations, the relative contribution to these value functions that reflect group versus self, leverage given to emotional versus empirical considerations, the value assigned to exploration, and so on. We base assessments of culpability, praiseworthiness, and moral responsibility for a decision on an agent’s policy.

To the libertarian, then, neural and environmental noise leads to a kind of indeterminacy that frees the philosopher from the shackles of determinism, but it does not confer responsibility because one cannot be responsible for undetermined acts. However, the way the agent (or agent’s brain) trades off speed and accuracy is an expression of that agent’s own approach to the noise. One setting may lead to more accurate decisions, which we may value in some circumstances or find ineffective in other circumstances when time is costly. Of course, like other policies, the speed–accuracy trade-off is controlled by machinery, and it is established through learning and experiences which are themselves shaped by the agent’s history of exposures and decisions. However, for reasons I will develop below, they cannot be accounted for by an unbroken chain of cause and effect that begins outside the brain of the agent.

Perceptual Decisions

The neurobiology of decision making has concentrated on relatively simple decisions concerning the relative value of one or another option or the relative merit of one hypothesis in relation to another. Often the latter

involves an interpretation of sensory data. Such perceptual decisions have proven tractable to neuroscience in part because of a strong foundation in perception, psychophysics, and sensory neuroscience (Green & Swets, 1966; Parker & Newsome, 1998). Perceptual decisions are based on evidence that is represented by neurons in the sensory cortex. This evidence guides a decision about some aspect of the stimulus. Roskies focuses on a well-studied example of perceptual decisions about motion. Here the experimenter supplies a sequence of noisy information in the form of a dynamic random-dot display. The subject's job is to decide if the net direction of motion is to the right or left, say. The decision process refers to the steps between the representation of evidence and the point of commitment to an answer (Gold & Shadlen, 2007), even if that commitment is merely provisional (Resulaj et al., 2009).

This simple perceptual decision has taught us much about how the brain accumulates evidence, works out the odds that one option is better than the other, and commits to the choice. It helps that so much is known about neurons in the visual cortex that sense motion and where those neurons project (Born & Bradley, 2005). It also helps that much is known about the way the brain controls eye movements (Schiller & Tehovnik, 2001). By training monkeys to indicate their decisions with an eye movement, it is possible to work back from the motor cortex to structures in the brain that receive the momentary evidence from the visual cortex. One of these areas is in the parietal cortex, termed the lateral intraparietal area (LIP) (Andersen & Buneo, 2002). The activity of these neurons shows us how evidence is converted into a commitment, that is, the neural mechanism of decision making.²

Like many neurons in the so-called association cortex, LIP has properties that allow it to accumulate information, to hold on to it, and to signal quantities based on information that has come and gone in the world. Moreover this activity does not precipitate a body movement. Thus cognitive functions like deliberating toward a decision are made possible by an elaboration (via evolution) on a basic sensorimotor design. The principal elaboration, which is shared by much of the association cortex, is persistent graded neural activity—a freedom from immediacy, as it were, from evanescent changes in the world and from the real-time engineering demands of controlling body musculature. As an aside, this “intentional framework” might interest readers with leanings toward embodied cognition (e.g., Shadlen et al., 2008).

One of the interesting features of perceptual decisions is that they are variable, especially when they are difficult. In the lab, it is possible to ask a decision maker to make many decisions about an identical visual

stimulus (the decision maker does not realize that the stimuli are identical, thereby removing the logical impetus to achieve consistency). This variability is attributed to variation in the state of neurons and in particular to the fact that their signals are noisy.³ Roskies points out that this noise breaks the Laplacian chain of cause and effect, in this case linking a visual stimulus to a choice. I would like to elaborate on this noise. Where does it come from, and what are its implications?

Neural Noise

Noise is a general property of cortical computation that can be appreciated from a simple consideration. Cortical neurons do not just receive and pass on spikes or turn on and off. They do something much richer with information. They compute with intensities (i.e., quantities, amounts, values) received from different sources to calculate some new intensity. These intensities are encoded by spike rates. This sounds simple enough, but spike rate is actually a more nuanced concept. Literally, spike rate is the number of action potentials (spikes) that a neuron produces per unit of time. Yet, for computation to occur quickly, the rate must be shared across neurons. Therefore spike rate is really the number of spikes from an ensemble of neurons per (brief) unit of time. The ensemble ensures that the intensity is represented continuously in time and thus available for computation in the silent gaps between the spikes of any one neuron. It allows neural computation to be fast, although there are limits (Mazurek & Shadlen, 2002).

There is nothing controversial about this concept of an ensemble rate code although it implies that each neuron receives input from many neurons. More to the point, for every spike a neuron emits, it probably receives 50 to hundreds of excitatory inputs from other neurons. This “high-input regime” is special to cortex (Shadlen & Newsome, 1994, 1998). It does not occur in simple nervous systems or in the brainstem of a mammal, and it is not even represented in the popular brain-slice preparation from the cortex itself (because most of the connections are not active or “sliced” off). The high-input regime does pose a significant challenge, however, because it implies that neurons receive a surfeit of excitatory input, which must be kept in balance with inhibition. Thus a generic function of the cortical microcircuit is to maintain a balance of excitation and inhibition (E-I balance, for short; Shu et al., 2003). Were it not for this E-I balance, cortical neurons would be unable to support graded spike rates. They would spike like crazy or not at all.

The same E-I balance explains the noise in neural computation. It does not explain it away but rather explains why it is present and ultimately irreducible. The gist of the argument is that spikes are no longer caused by a few excitatory events but rather by a process likened to threshold crossings of a membrane voltage that drifts as in Brownian motion (Gerstein & Mandelbrot, 1964). The Brownian drift is the result of the barrage of excitatory and inhibitory synaptic events. The times between threshold crossings (also known as first passage times) would exhibit tremendous irregularity, just like the intervals between spikes of cortical neurons.

In her essay, Roskies expresses agnosticism over whether the irregularity of cortical spiking is truly noise or just unexplained variance. I do not share this agnosticism. There are both principled and empirical reasons to believe that the irregularity should be viewed as noise. The gist of the principled argument goes something like this. The time of the spike is determined by the sequence of E and I events that preceded it, but to view the time of a spike as informative would require that in some sense the events that caused its timing could be reconstructed. This is highly unlikely. While it is true that the exact same sequence of E and I events would lead to a spike within ± 1 milliseconds with high reliability, the number of possible sequences that could lead to this same well-timed spike is enormous. This is just another side of the Brownian motion comparison. There are many paths that lead to the same threshold crossing. This information is lost, however, because the history of the voltage of the nerve cell's membrane is not passed on to other neurons. Only the resulting spike survives. This argument requires more formal justification, but it has led theorists to regard the balanced E-I, high-input regime as chaotic (van Vreeswijk & Sompolinsky, 1996). Importantly, the observation that a neuron would respond nearly identically to the same injected current (Mainen & Sejnowski, 1995) is irrelevant.⁴

The empirical argument is simply this. The irregularity explains variation in choice: the overall error rate, trial to trial variation in reaction time, and confidence in the decision. Although this does not prove that the variability is unexplained, it does imply that such variability is unknown to the higher brain areas that convert evidence into a decision. The most convincing cases in which neural activity can be related to decisions—in some cases causally—are those in which the variability impacts the decision as noise (Parker & Newsome, 1998). If the variation were caused by some irrelevant feature of the environment or the agent's internal state, then the decision mechanism should know to discount this variation—that is, to explain it away. This is something the brain does routinely with

changes in light level, volume, confounders, and nuisance parameters. Yet, when it comes to decisions, even simple ones, the brain makes errors because a neuron (or ensemble of neurons) emitted a few spikes more or less. In other words, the decision cannot cleave the signal from the noise.⁵

To summarize, noise is fundamental to cortical processing and hence cognition. It arises naturally as a by-product of computation in the high-input regime that supports cortical computation, and it may therefore be viewed as beneficial. As Roskies points out, noise does not confer a basis for freedom or responsibility. Yet it solicits from the agent (or agent's brain) a policy on whether and by how much to tame this noise. The neural mechanisms supporting bounded evidence accumulation support just such a policy. It is upon such policies that we hinge notions of responsibility, regret, and other attributes that arise in ethical dialogue.

I will next explain why both noise and policy settings invite us to reject the form of determinism that arises in consideration of free will and responsibility.

The Falsity of Determinism: Chaos and Emergence

Roskies states, "We have no objective access to either determinism or indeterminism." I don't know why not. It is beyond my expertise to weigh in on the philosophical distinction between metaphysical and apparent determinism, but I suspect it is problematic mainly because of what is meant by a state. Consider that some mechanical systems, which obey lawful dynamics, can exhibit chaotic behavior. I believe we are too quick to write off the apparent indeterminacy of a chaotic process to "epistemic" limitations. The notion of a state with an attribute of infinitesimal variance is itself an abstraction based on a concept that was invented for the calculus—an idealization, albeit a useful one. In many instances, it is not that we cannot measure states with sufficient precision but that precise state is itself chimerical. Rather, it is that we do not live in a universe in which there is such a thing as an infinitely precise definition of state.⁶

The lesson from the simple mechanical systems used to illustrate chaos is not that chaos is actually deterministic but that it is much easier to break determinism than it would seem. If there is any unaccounted-for variation, it will grow at such a rate as to refute the conditions necessary for determinism. No quantum mechanics required for assembly.⁷

Here is another way to think about this. Depending on the system we are scrutinizing, entities interact and states evolve in ways that tolerate such finite variation. This allows us to infer regularities in physics and

chemistry, as in A interacts with B to produce complex AB. The reaction occurs despite a range of variation in the states of A and B. Often state AB retains less variation than the states of A and B that led to its complex. The lost entropy is at the expense of energy, and it implies a loss of information about A and B preceding their union because its record is limited to the finite variation in the state of AB. In some cases the variation that is passed on to AB grows, and depending on the process, it might even grow exponentially, in which case we refer to the system as chaotic. Both cases are relevant to the neurobiology of decision making. The first, information loss, is relevant to the understanding of the neural noise mentioned in Roskies's essay. The second, expansion of entropy, implies indeterminism is a general property of our universe (e.g., the second law of thermodynamics) and in particular that a form of radical reductionism that would explain the state of the decision maker as the result of external causes will often fail. Here is how these considerations interact with the relevant neuroscience.

The main implication of information loss is that the time of a spike cannot convey which of the many possible paths led to it. This is especially true if we accept that there is a finite Δt around the time of the spike that would be indistinguishable in the sense that future states of the neuron and any neuron to which it projects are unable to tell the difference. The magnitude of Δt is an empirical matter for neuroscience although it is currently unknown (half a millisecond is a safe guess). If the assumptions of the high-input regime hold, one could argue that precise spike times (meaning within this same Δt) are unlikely to convey information in cortex. Precise spike times govern the particular trajectory taken by the neuron's membrane voltage between spikes, but no other neuron in the brain has access to this information. Yet, if we accept that the balanced E-I leads to sensitivity to starting conditions (and exponential divergence), we can appreciate why world events such as turning on a random-dot motion display will catch the brain in a random state and therefore why identical stimuli lead to variation in the timing of spikes (and the loss of the information that would permit reconstruction, even in principle, of the causal history leading to the spike time). This variability in timing of spikes translates to variation in the number of spikes present in an epoch from a single neuron or from an ensemble. This is the basis of the neural noise that limits the fidelity of perception, the precision of movement, and the speed versus accuracy of decisions.

These considerations mark a break from reductionism without abandonment of causal explanation or mechanism. I suspect that there are many

other instances in biology, physics, and the social sciences which exhibit similar “breaks,” as they may be characterized as follows. A chain of cause and effect led from state A to endpoint X; yet (1) given A and complete knowledge of the mechanisms at play, it is not possible to predict X with certainty, and (2) X does not retain a complete or useful record of the chain from A. This recipe for emergence has been harnessed against radical reductionism by the evolutionary biologist Ernst Mayr (Mayr, 2004) and the physicist P. W. Anderson (Anderson, 1972). The common feature is an expansion of the possible paths forward from an initial state and a collapse of state (e.g., to a spike or compound AB or species or phase transition) that marks a loss of the information that would be required to retain the details of the causal history and transmit them as the path evolves further.

Of course, unreconstructability is compatible with determinism, but in combination with (apparent) indeterminism, it supports a type of emergence that is relevant to considerations of responsibility. It implies that it is fruitless—not just difficult but impossible—to find a more ultimate explanation for the state or property at the point of the chain under consideration. This is not an abandonment of mechanism; it is a critique of radical (eliminative) reductionism. Its relevance to the argument at hand is that policies upon which we hinge responsibility must be reckoned with at the level of the agent, which I take as equivalent to the agent’s brain. This is what distinguishes the faulty/laudatory decisions of an agent from those of a machine. For the latter, we feel it is sensible to explain a fault/attribute with considerations wholly outside the machine. We do not punish or praise the machine the way we do an animal or a human.⁸

These considerations apply more broadly to the latter sections of Roskies’s essay concerning mechanism and reductionism. She states, “The reductionist worry is that neuroscience will show that behavior can be reduced to laws operating at a nonpsychological level.” I agree with Roskies that this outcome seems unlikely. This is because the chain of cause and effect contains breaks of the type that undermine radical reductionism and determinism,⁹ at least in the form required to undermine freedom (per incompatibilism). Perhaps this is what she means by a rehabilitated concept of causation: “It is possible that a rehabilitated conception of causation could alter the concept of mechanism sufficiently to make it more congenial to compatibilist views.”

The argument against reductionism is especially germane to the consideration of those “policies” that govern decision making and upon which we attach moral assessment of responsibility, regret, and praise. Policies are controlled by neurobiology with settings established by the organism’s

history, but it is not possible to provide an exhaustive explanation of these settings by looking at the genome, childhood experiences, and so on. In this sense, policies resemble Kane's self-forming actions (Kane, 2002). These and other libertarian constructions seem consistent with the neurobiology and the argument against eliminative reductionism sketched above. This is why I find Roskies's essay more interesting to libertarians than she probably intended.

Closing Remarks

As a neuroscientist concentrating on the problem of decision making, I am sympathetic to Roskies's view that neuroscience has something to contribute to philosophy, although I think our role is restricted. I disagree with colleagues who would give neurobiology a seat at the head of the table. I am content to sit at the children's table or to be invited in to provide the entertainment. Here, I have attempted to elaborate a richer view of neural noise and its role in perceptual decisions. I have suggested ways that it might bear on long-standing debates about determinism and free will, but I am content to leave this to the experts.

When it comes to the practice of ethics, however, neurobiology may have more to say. If there is merit to the argument about policy—that this is a neural level of organization that bears on considerations of responsibility—then neuroscience and neurology may play a more significant role. Knowledge of the neural mechanisms, their dysfunction in disease, and their maturation in the developing child or adolescent could one day affect considerations of blame, punishment, and normative expectations.

One interesting consideration is the necessity of consciousness for underpinning moral responsibility. If neurobiology shows that the mechanisms underlying complex decisions operate with and without conscious awareness, then it seems to me that we might hold agents responsible for some decisions even if they are made unconsciously (without impairment). An emerging consensus in both neurology and neuroscience is that much cognition occurs without the aid of consciousness. There are fascinating questions that arise about how unconscious and conscious processes interact. I have written that the neurobiology of conscious and nonconscious cognitive functions probably utilize similar mechanisms. Both involve decisions, and in particular decisions to engage in certain ways (Shadlen & Kiani, 2011). When those ways involve the possibility of navigating, planning to (possibly) reach, look, grasp, and the like, they are implicated in spatial awareness. When the ways involve the possibility of pointing

out, planning to (possibly) report, make narrative, or recall later for purposes of communication (even with the self), then they support much of what we think of as consciousness.¹⁰ The main virtue of this perspective is not that it provides a magical solution to the vexing problems of subjectivity or language.¹¹ Rather, it suggests the neurobiology will probably not need any more magic here than it needs to understand some of the capacities that seem less mysterious, like the simple decisions that animals make in the laboratory.

I suspect that the neurobiology of consciousness will guide our concepts of free will, because the latter hinges on both volition and subjective awareness of volition, what is termed authorship. However, Roskies advises the field “to explore the possibility that freedom is a concept derivative on more robust intuitions about responsibility rather than vice versa.” For responsibility, neurobiology helps to expose why different agents make different decisions, and these considerations, including policies, do not require subjective first-person reports. To engage in ethics—to explain, justify, condemn, praise, and punish—requires not only consciousness but the actual dialogue. An exchange between philosophers and neuroscientists is a step in the right direction.

Notes

1. Roskies and I have argued elsewhere (Shadlen & Roskies, 2012) that these policies represent a level of organization—in behavior and neuroscience—that connect neuroscience to matters that interest ethicists and philosophers. The trade-off between speed and accuracy is the best understood. In short, neurons in the association cortex accumulate evidence sequentially, when it is available, in a process that resembles deliberation (n.b., consciousness not needed). This integration improves the strength of the signal in relation to the noise, but at some point other neurons sense that the level of activity is sufficient to stop accumulating evidence and to terminate the decision process. This level represents a policy (or attitude or value) concerning the relative merit of accuracy versus speed.
2. LIP is not the brain’s decision center—I wonder if such a center exists—nor is decision making LIP’s sole function. LIP plays a role in the decision process that is reduplicated by other areas of the association cortex. Importantly, LIP illuminates neural mechanisms that are shared by many brain areas which are involved in decision making and other cognitive processes.
3. In psychophysics and neuroscience there seems to be no place for so-called “torn decisions.” Even guessing is supported by the noisy evidence, which favors one or the other alternative. When a subject is performing at chance, it is not because the

evidence is at chance; it is because it is equally likely to be on either side of chance on any one decision. Nor can we trust subjects' intuitive report that they are only guessing. These reports correspond to confidence, but they do not correspond to actual performance, which can be substantially better than chance, despite such a report.

4. I disagree with Roskies's assertion that this study "suggests, but falls short of demonstrating, that neurons do not fire stochastically to biologically realistic inputs but rather respond with action potentials in highly predictable, stereotyped ways." It merely rules out the spike threshold mechanism as an explanation for the observed variability *in vivo*. (I agree with the "falls short of demonstrating" bit.)

5. It also does not make errors by matching probabilities. If the neurons compute that option A is better than B with probability 0.7, it does not consult a random number generator and match this probability. It decides/answers "A." It is a common mistake to attribute stochastic variation in behavior to a process that matches a probability. The stochastic variation (across repetitions) arises by selection of the best option in each instance. The variation is explained by signal-to-noise considerations on an otherwise deterministic mechanism. When the brain decides "B," it is because the evidence plus costs favors B. Of course, costs associated with exploration and perseveration can support a choice that trumps the evidence. But that should not be confused with matching a probability.

6. The issue boils down to whether a single sample of "state" from its distribution (e.g., in statistical thermodynamics) can be regarded as existentially precise. When we consider variation in state, we are thinking of an ensemble of molecules or repetitions and we are content to represent state as a probability distribution. But what about the particular state—the single sample? Is it really sensible to conceive of it as possessing no variance, or is it better represented as a probability distribution (i.e., possessing finite variance)? In this sense, the issue parallels an old debate in the theory of probability. The frequentist objection to the Bayesians was not about the arbitrariness of priors but about whether an actual state of the world (or hypothesis) should be regarded as a probability (Howie, 2002).

7. Quantum mechanics may come to play in a deeper consideration (beyond my understanding) of the second law of thermodynamics.

8. For why animals are not ethical agents, see Williams's (1985) comments on speciesism (p. 118).

9. Again, I am open to the possibility that (1) the form of indeterminism I am supporting is the apparent type, in which case the thesis (like Roskies's) is that this is what matters, or (2) this indeterminism has metaphysical status because the concept of metaphysical indeterminism (MI) rests on a faulty definition of state, and/or (3) the distinction between MI and apparent indeterminism demands further scrutiny.

10. A provisional plan to report to another agent extends perception beyond affordances. Locations are no longer merely places to approach, avoid, look at, and reach for, and objects are no longer merely things we consider eating, mating, and exploring. The possibility of reporting to another gives the place and object a presence in the world shared by another agent with a mind like mine. And memories do not simply call up what we might next do—as in embodiment or Gibsonian affordances—but something more like narrative. The possibility of reporting to another agent with a mind like mine extends to a report to myself in the future or reevaluation of what my own mind would have experienced in the past. Decisions to possibly report naturally introduce subjective, episodic content with features that mix *res extensa* with *res cogitans*, a view expressed by Merleau-Ponty (1962).

11. In neurobiology, these magical solutions are associated with oscillations, synchrony, resonance, and nonlinear dynamical systems.