12

Robustness in a Variable Environment

Kevin A. Gluck, John M. McNamara, Henry Brighton,
Peter Dayan, Yaakov Kareev, Jens Krause,
Robert Kurzban, Reinhard Selten, Jeffrey R. Stevens,
Bernhard Voelkl, and William C. Wimsatt

Abstract

Robustness is a prominent concept in technical sciences and engineering. It has also been recognized as an important principle in evolutionary biology. In this chapter, it is proposed that the term "robustness" be used to characterize the extent to which a natural or artificial system can maintain its function when facing perturbation and that this concept is relevant in considerations of Darwinian decision theory. Situations in which the action of natural selection is liable to lead to the evolution of robust behavioral strategies are highlighted along with some psychological mechanisms that might lead to robust decision-making processes. Robustness describes a property of a system varying on a continuous scale rather than existing as a dichotomous feature. Degree of robustness depends on the details of the interaction of system characteristics and environmental contingencies, as well as the specific types and extents of perturbations to which the system may be subjected. A system can be robust in one domain while remaining highly vulnerable to perturbations in others. As defined here, robustness is related to, yet distinct from, flexibility and optimality. The sorts of environmental variation, and hence perturbations, that an organism or technology is liable to face are described, as is the cost-benefit trade-off of robustness. Finally, the robustness of decision making at the level of social groups is considered.

Introduction

Soldiers fighting in the trenches during World War I reportedly developed a "live and let live" mentality in which mutual truces spontaneously arose, but a violation of the truce resulted in quick retaliation. Axelrod (1984) proposed this as an example of tit-for-tat, a strategy that can maintain cooperation by copying a partner's previous action. In the trench warfare example, ceasefires

From "Evolution and the Mechanisms of Decision Making," edited by Peter Hammerstein and Jeffrey R. Stevens.
2012. Strüngmann Forum Report, vol. 11, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01808-1.

led to ceasefires and unprovoked attacks led to retribution. Though tit-for-tat can generate cooperation, a simple mistake can spawn a spiral of escalating aggression. An itchy trigger finger can start a cascade of violent attacks in the trenches. A forgotten or misplaced thank you note can explode into a dramatic family feud. As a decision strategy, tit-for-tat may work well in a perfect world with no mistakes or accidents, yet it is not *robust* to error in the environment (Selten and Hammerstein 1984).

The topic of robustness is integral to the evolution of biological organisms (Hammerstein et al. 2006; Kitano 2004; Wagner 2005). The same is true for the evolution of nonbiological systems, such as scientific communities and technologies. Because decision makers' environments are and have been subject to perturbation, robustness is important to consider in the context of a general theory of decision making, and especially so in a decision theory built on the principles of Darwinian evolution.

Examples of robust solutions to problems posed by perturbations include, but are not limited to, the existence of multiple (often redundant) mechanisms, tolerance buffering, and systemic stochasticity. The competition and cooperation among these and other mechanisms in complex and uncertain environments give rise to the observed complexity of decision-making behavior (Simon 1996).

Our main argument is that robustness has been a key driver for the evolution of different decision-making and control mechanisms. Choice reflects the competition and cooperation of these mechanisms and therefore does not fit nicely into categories which arise from a simple set of axioms. Much hinges on the necessity of learning, which is a central battle between flexibility and speed of response.

Understanding the evolutionary origins of decision mechanisms requires us to address the nature of robustness, the nature of environments, and the costs and potential benefits associated with robustness in these environments. Acknowledging that robustness can result from social interactions, we also consider robustness in group decision making, especially via emerging results from the literature on swarm intelligence.

Robustness

The concept of *robustness* is important to many fields of science. In structural engineering, structures must meet a fundamental requirement of robustness. For example, as described in structural design codes EN 1990 and EN 1991-1-7, "a structure shall be designed and executed in such a way that it will not be damaged by events such as explosion, impact, and/or the consequences of human errors, to an extent disproportionate to the original cause." In biology, robustness addresses the probability of survival and reproduction for members of a given species. From economies to ecologies, from *Homo sapiens* to

Drosophila, from the physical landscape in which animals forage to the conceptual landscape in which scientists forage, and across all spatial and temporal levels of analysis, robustness matters. The broad relevance and complexity of the concept of robustness prompt us, however, to ask: Is there a level of description of robustness that can aid our understanding of decision making?

Defining Robustness

To begin, we must clarify our perspective by defining robustness and some of its features. Robustness can be conceptualized as a function between some dimension of the environment and performance. For example, for a given automobile engine, there is a relationship between temperature and performance such that performance deteriorates at low and high temperatures. Different engines will exhibit different curves illustrating this relationship. Such curves can be summarized in various ways (e.g., mean, standard deviation, area under the curve) to characterize how "robust" the engine is *with respect to* temperature. Greater performance across the environmental dimension equates to greater robustness. A selling point of the original air-cooled Volkswagen engine, for example, was that it would actually start (at least more reliably) at colder temperatures than its water-cooled competitors. Within the Volkswagen engine, a key performance characteristic (startability) was more robust to temperature variation than the other engines on the market at the time.

In the context of biological systems, different phenotypes can be understood to have different performance robustness curves—how well the system in question executes its evolved function—under a range of different conditions. Over evolutionary time, mechanisms will be selected whose properties maximize reproductive success. The specific mechanisms selected by Darwinian evolutionary processes depend strongly on the frequency and magnitude with which different environmental contingencies are encountered and on the implications of the interaction of biological mechanism and environmental contingency for reproductive success.

In the context of computations, one can construe performance as a relationship: the production of a particular output (e.g., representation, behavior) given a particular state of the world. The same analysis applies. Different mechanisms will maintain performance—this adaptive, systematic input-output relationship—to a greater or lesser extent, and we can try to quantify how "robust" each mechanism is: How well does the mechanism maintain its input-output relationship as a function of some environmental parameter? We define robustness as:

> *the extent to which a system is able to maintain its function when some aspect of the system is subject to perturbation.*

Robustness is not a binary category but rather a continuously varying *n*-dimensional state space. This conceptualization of robustness is consistent with

From "Evolution and the Mechanisms of Decision Making," edited by Peter Hammerstein and Jeffrey R. Stevens. 2012. Strüngmann Forum Report, vol. 11, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01808-1.

Levins's (1966) emphasis on *invariance*, Campbell's (1958, 1966) emphasis on *multiple determination*, and Wimsatt's (1981) methodological prescription for *robustness analyses*. High dimensionality makes quantifying robustness in a general way more challenging and less tractable but not necessarily impossible, provided one can formally specify the dimensions and possible perturbations. Systems, then, will exhibit a "degree of robustness" as opposed to being dichotomously robust or nonrobust.

We are using the term *system* here in the most general sense, and an aspect of that system is any variable, component, characteristic, or policy that could affect its function. We adopt a similarly broad interpretation of *function*: it encompasses a system's purpose, process, or level of performance or outcome. A *perturbation* is any pressure or stressor or environmental change that could in principle affect the system's function.

Flexibility, Optimality, and Robustness

Robustness is associated with *flexibility* and *optimality*. Though related, these concepts are not identical to robustness. Confusion surrounding the relationships between and distinctions among these terms seems to originate in the fact that flexibility, optimality, and robustness are all potential capacities or characteristics of systems. The difference is that robustness depends on, as defined earlier, *the maintenance of function when some aspect of the system is subject to perturbation*. Thus, robustness is only identifiable within the context of environmental variation and perturbation, and is only present when the system's function is preserved. In other words, in a static environment it may be possible to demonstrate that a particular system is flexible or even optimal (when the environment is both static and certain) within whatever fixed dimensions define the operating environment, but this tells us nothing about the system's robustness. Robustness is only demonstrable when the function of the system is assessed across perturbations in environmental contingencies.

Figure 12.1 provides an abstract illustration of some of the key issues for robustness and allows us to distinguish it from optimality (defined according to some performance criterion) and flexibility (in terms of adapting to prevailing conditions). The figure is intentionally abstract to make general points about systems and environments and the influence of interactions of their characteristics on functional performance.

For expository purposes, let us make consideration of the abstractions in Figure 12.1 more concrete by assuming that the figure represents the performance of a hypothetical agent foraging in an environment, which can potentially exist in different states, modeled as a scalar value $e \in [0, 10]$. This could represent some characteristic of the environment, such as the rate at which prey arrive. Similarly, the agent's decision policy is parameterized by a scalar parameter $p \in [0, 10]$ (say a measure of its preference for exploitation rather than exploration). Figure 12.1a shows the performance of the agent as a function of

From "Evolution and the Mechanisms of Decision Making," edited by Peter Hammerstein and Jeffrey R. Stevens. 2012. Strüngmann Forum Report, vol. 11, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01808-1.

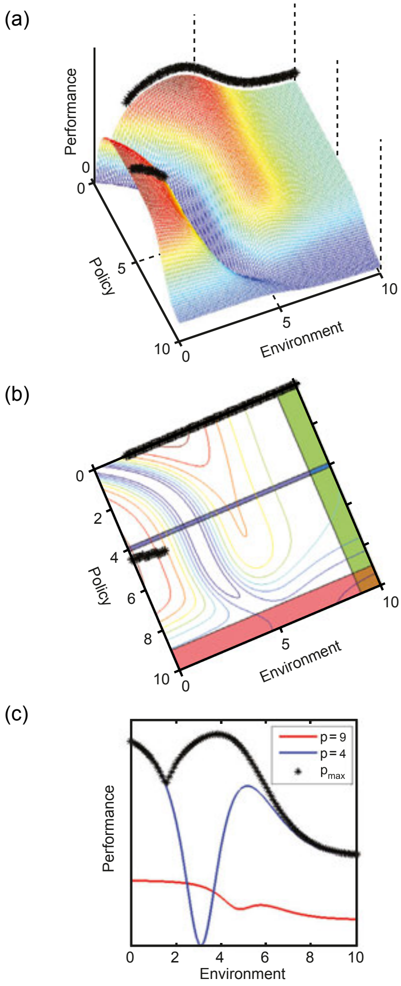

Figure 12.1 Issues surrounding robustness: (a) and (b) illustrate a hypothetical case in which the performance of a system depends on a single scalar quantity representing the external environment (designated *e* in the text) and a single scalar policy (designated *p* in the text) representing an aspect of the system. A 3D plot of the performance surface that results from interactions of the policy-environment interactions is shown in (a); a contour plot of the same surface is provided in (b). The black points show the optimum value of the surface for each value of the environment for the parameter range shown. The colored rectangles on (b) refer to ranges of values of the environment or the policy whose performance characteristics we highlight in the text. In (c), performance is shown as a function of the value of the environment (*e*) for various policy values in (b), as described in the text.

e and *p*, which, for the moment, we will consider to be measured over a short time period. Figure 12.1b shows the same surface, but now as a contour plot.

Let us further assume that in this context the function of the agent is to maximize its foraging performance. We can now consider how the performance of the system varies over the interaction of environment and agent characteristics:

1. If the environment only occupies a limited range of values of e, say $e \cong 2$, then an agent with a fixed policy (here with $p \cong 4$) will optimize performance.
2. However, if the environment changes a little, so that $e \cong 3$, now performance with $e \cong 4$ is catastrophically bad. Thus, we might say that performance with a fixed parameter p has low robustness to environmental perturbation. The blue curve in Figure 12.1c shows the performance as a function of e for $p = 4$; the catastrophe at $e \cong 3$ is apparent.
3. If, on the other hand, we choose $p = 9$ (the red curve in Figure 12.1c), then performance is more stable as a function of e. Maintenance of performance, in terms of minimizing performance variability, is one popular measure of robustness; note, however, that this comes at the cost of rather poor performance in environments with low values of e compared with what could possibly be achieved for those environments. Indeed, a credible strategy for explaining apparent sub-optimalities in the behavior of an agent in some environment is to declare that this provides robustness in other possible, but not current, environments.
4. Conversely, if the environment happens to take the value $e \cong 10$ (the green rectangle in Figure 12.1b), then performance is very insensitive to the agent's policy. This is another form of robustness; that is, if the policy is subject to internal perturbation arising from damage or decay.
5. A further possibility is that the agent might be able to measure the state of the environment and adapt its parameter p accordingly. The black asterisks in all of the subplots in Figure 12.1 show what happens if the agent can choose p to optimize performance as a function of e. Performance is attractively good. This shows how flexibility can aid robustness. However, this flexibility comes with three potential costs: two statistical and one computational. The first statistical cost stems from the need to measure the value of e. This requires samples, which could be expensive to acquire (if the values of p are inappropriate during this period). These costs can be mitigated by knowledge of the likely environments e, which will help constrain the estimation problem. The second statistical cost is that of estimation error/uncertainty. Even if estimation is performed as well as possible, there is a chance of unlucky samples that could skew the value of e that is inferred, giving rise to poor performance. The computational cost comes from the extra machinery necessary to make the agent exhibit this flexibility.
6. For the case of the blue curve, even a small change in e suffices to ruin performance if that small change results in $e \cong 3$. Another popular notion of robustness concerns outliers (see Dayan, this volume). What

happens if the environment can sometimes take the value $e \cong 20$, which is outside any previous experience? Human scientists and engineers can use theories about the world to extrapolate surfaces, such as the one in Figure 12.1, beyond any extant observation, and so protect against imaginable but not experienced environments. Natural selection does not have this luxury. Indeed, as we discuss below, natural selection is condemned to respond to the history of recent environments, and thus can exhibit a form of maladaptive and non-robust overfitting.

7. Finally, we face the necessary task of integrating the net performance curves, such as those in Figure 12.1c, in order to assess the total performance of a particular policy. For instance, it might be catastrophic for performance ever to go to 0, or maybe only the average performance over the distribution of possible environments matters. Unfortunately, this step is completely dependent on the range and distribution over possible environments, and also the problem. Even for the case of natural selection, it will be different if we measure the expected number of offspring at different times in the future, or the probability of not having been eliminated. Thus, whether we consider the red curve in Figure 12.1c to be better or worse than the blue curve is formally unanswerable in any absolute sense. The answer is contingent on the specific circumstances.

In summary, there are clear circumstances under which we can distinguish robustness from at least a local notion of optimality and from flexibility. It is less possible to distinguish robustness from a more holistic notion of optimality, since the latter can always be structured so as to include robustness.

It is sometimes supposed that there are trade-offs between the plasticity or flexibility of behavior and its robustness. In such circumstances one must first ask whether the correct parameter is being evaluated. Thus flexibility of behavior may contribute to the robustness of a fitness outcome. With the robustness of a fitness outcome, would there then be no pressure for evolutionary change, restoring the opposition between robustness and flexibility?

At other times, robustness is pitted against optimality in discussions of decision making (e.g., Rosenhead et al. 1972). Arguments are usually about whether the optimal solution for a model is robust in a wider modeling setting or in the real world. Of course, in the real world, it is not possible to identify with certainty what is optimal (or what is robust) because it is not possible to specify fully the real world in modeling terms. Nevertheless, assuming that what does evolve is adaptive, this argument can be loosely reformulated as: Is it adaptive to be robust, or is it optimal to be robust?

Here we focus on the implications of robustness across evolutionary environments for the case of decision making. In some cases (such as inbuilt fight, flight, or freezing responses), the agent's behavior appears to be fixed and genetically determined, perhaps as a response to the dangerous costs of

From "Evolution and the Mechanisms of Decision Making," edited by Peter Hammerstein and Jeffrey R. Stevens. 2012. Strüngmann Forum Report, vol. 11, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01808-1.

learning. In other cases agents achieve good performance across a range of environments (which we consider a form of robustness) by learning (Dayan, this volume). However, there are multiple mechanisms for learning, which, again, may putatively be a response to the demands of performance across variable environments.

Variable Environments

If everything about the environment in which a system exists is static, then robustness is irrelevant. In the real world, of course, environmental variability is ubiquitous. It is not a question of *whether* there is environmental variability, but rather *what, when, where*, and *how much* variability exists. In our definition of robustness, we refer to these environmental variations generically as perturbations.

The notion of robustness has broad application because operating conditions are rarely stable. Assessing a system's robustness requires specifying a range of environmental perturbations over which the system will to a greater or lesser extent maintain its function. Yet what forms of variation do organisms commonly encounter? In particular, by categorizing common forms of variation, is it possible to shed light on robust design patterns? Before considering some key dimensions of environmental variation, it is worth stressing that the list of potential sources of variation is unbounded. Furthermore, the relevance of a particular form of variation will depend both on the organism or technology in question and the level at which we seek to understand it. A cognitive scientist, for example, will likely find irrelevant the fact that all biological systems operating at temperatures above 0 Kelvin face the perturbations arising from thermal noise, even though such perturbations place constraints on functional design (Wagner 2005). For the purposes of this discussion, we wish to describe a generally useful categorization of dimensions of environmental variability relevant to robustness:

1. *Variation within and between environments*: Variation occurs both within and between environments. For instance, within a specific environment, the relevant operating conditions faced by an organism could be relatively stable, such as the temperature and light conditions of deep-sea dwelling creatures. However, the range of potential environments that organisms of the same species may face could be highly variable and uncertain. Consequently, the level at which we examine variation (e.g., the population vs. the individual) will determine which forms of variation are relevant to robustness (Wagner 2005). Indeed, the following forms of variation have the potential to occur both within and between environments.
2. *Internal versus external variation*: Variation can refer to, say, changes in the state of an organism's endocrine system or its energy levels.

Should such factors be considered part of the environment or part of the organism? Intuitively, the body defines a boundary separating the internal and external environment, both of which present sources of variation which could pose a threat to functioning. Faced with the problem of regulating its own temperature, for instance, an organism will be subject to both internal and external determinants of temperature change. Robust responses to these factors may require different designs, due to internal and external variation occurring over different timescales (Wagner 2005). More generally, fatigue, aging, and a developing immune system are all examples of variation in the internal environment which place constraints on functional design.

3. *Degree of predictability*: Uncertainty exists when an action, such as a movement, a decision, or an interpretation, has more than one potential outcome. Variability in these outcomes can be more or less predictable; that is, cues available to an organism vary in how well they correlate with the true state of the world. For example, distinguishing predator from prey, moving from one food patch to another, or interpreting a visual scene all require inferring latent properties of the environment from environment cues. In each case, the available cues provide more or less predictive indicators of events. At one extreme, all mechanisms are equally robust in a completely random, maximally uncertain environment. The presence of regularities, however, will mean that mechanisms will be robust over certain ranges of variation at the expense of others (Geman et al. 1992). Thus, a key dimension of all forms of variation, and one which strongly influences robust design, is the degree of predictability.
4. *Degree of stationarity*: Some forms of variation, such as the caloric content of alternative food items, will likely remain constant over time. They are *stationary*, in that the statistical properties (e.g., mean, variance) of the caloric content of specific foods remain constant over time. Other forms of variation, such as the where the food is located, will likely change over time. They are *nonstationary* in some combination of time and space. On the one hand, a temporal or spatial dependency can be seen as another cue and simply an additional source of uncertainty which renders events predictable to a greater or lesser extent. On the other hand, both the prevalence and significance of nonstationary properties of environments suggests that they are worthy of study in their own right. For example, the balance struck by an organism when facing the trade-off between exploration and exploitation will likely depend on how, and to what extent, the environment changes over time and place (e.g., McNamara and Houston 1985b). In this way, it is not only environmental variability but also variability (nonstationarity) in environmental variability that is a dimension on which it is relevant to consider robustness.

5. *Social determinants of variation*: Social variation arises from the actions of other agents, rather than, say, nonsocial variation (e.g., the caloric content of food items or seasonal variation). Although the dividing line between social and nonsocial environments can often be unclear, the social environment nevertheless introduces forms of variation which would otherwise be absent (Hertwig et al. 2012), and therefore it is also germane to considerations of robustness. Strategic, game-theoretic settings represent one class of variation in which other individuals' behaviors depend on the agent's own actions. This contingency can result in complex and dynamic variation in the social environment. The human linguistic environment is a striking example, as the generative nature of human language continually leads to the creation of novel utterances and the evolution of interpretation of meaning (Brighton et al. 2005). Just as the consumption of food and water resources by others will influence the hunting and gathering decisions of organisms, the communicative acts of others can influence interpretation, understanding, and decision making.

The existence of multiple types of environment, affording different opportunities, allows for the possibility of modeling the environment. This can be done by setting hyperparameters in parameterized algorithms (i.e., parameters that describe the possible distribution of other lower-level model parameters which may, in turn, describe distributions of resources within the environment) or by employing preprogrammed modules when these have been hard-wired. The trouble with the latter is choosing between modules and organizing learning within the modules. Modules may or may not compete; an example of the latter is when modules process different information about the same underlying quantities. Generic strategies may exist to cope with these opportunities; for instance, apparent over-optimism (exploration bonuses; Kakade and Dayan 2002) in the face of ignorance, followed by over-caution in the case of disaster. Robust exploration methods, such as regret bounds for bandit problems (Auer 2003), are another possibility, with attractive, though sometimes suboptimal, characteristics. Representational learning, such as prolonging development so that the representations of the world on which decisions hang can be adapted to the prevailing statistical properties (see Dayan and Abbott 2001, chapter 10), is a further critical form of robustness.

Evolutionary Selection Pressures

There would be no evolutionary selection pressures to produce robustness unless population members experienced environmental perturbations. Thus we might expect the robustness of current decision mechanisms to be positively correlated with the degree and dimensionality of ancestral perturbations.

However, even when there are perturbations, whether we expect robustness to evolve depends on the particular perturbations and the performance measure considered, as the following examples illustrate.

Case 1: Amount of Food Obtained Each Day during the Winter

Consider a small bird or mammal that is trying to survive the winter. This animal has a very limited ability to store food as fat, and so must get enough food each day or it will die. Suppose that food availability varies from day to day. Then we might measure the performance of a foraging rule that the animal might use as the amount of food found for each level of food availability when using this rule. Some rules may do well on average, some may be especially good when food is plentiful, and some may be very efficient at finding food when food is scarce. Unless a rule has reasonable performance on each day during the winter, the animal will not survive the winter. Thus we would expect natural selection to produce foraging rules that maintain the capacity to find sufficient food each day, despite adverse environmental perturbations (i.e., we expect the evolution of rules that are robust given our performance measure).

Case 2: Number of Surviving Offspring Produced in a Breeding Attempt (Demographic Stochasticity)

Consider a hypothetical female bird that breeds once a year. During breeding she decides whether to lay one or two eggs. After the eggs hatch, the chicks tend to attract predators through their call. Thus, the more eggs the female lays, the greater the probability that a predator will find the nest and kill all of the offspring. Suppose that a predator may be present in the local area or not; each event will occur with a fixed probability. If no predator is present, then each egg laid results in a mature offspring. If a predator is present and the female lays one egg, then the predator finds the nest with a low probability. If she lays two eggs, the predator finds the nest with a higher probability. We assume that whether this female's eggs survive is independent of what happens to other breeding individuals (demographic stochasticity). Here our performance measure is the number of surviving eggs. The environmental conditions are whether a predator is present. Table 12.1 gives a numerical example.

In this setting we expect natural selection to favor the egg-laying strategy that maximizes the expected (average) number of surviving offspring. Suppose that the predator is present with probability 0.5. Then laying one egg will result in $(0.9 + 1)/2 = 0.95$ surviving offspring on average; laying two eggs will result in $(0.2 + 2)/2 = 1.1$ surviving offspring on average. Thus the laying of two eggs will be favored by selection. Note, however, that this strategy is not robust in that few offspring survive when the predator is present.

From "Evolution and the Mechanisms of Decision Making," edited by Peter Hammerstein and Jeffrey R. Stevens.
2012. Strüngmann Forum Report, vol. 11, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01808-1.

Table 12.1 Number of surviving offspring per year in nests with one and two eggs depending on whether a predator is present. The probability a predator detects a nest with one egg is 0.1; the probability that it detects a nest with two eggs is 0.9. The probability that a predator is present is 0.5.

	One egg	Two eggs
Predator present	0.90	0.20
Predator absent	1.00	2.00
Average over conditions	0.95	1.10

Case 3: Numbers of Surviving Offspring (Environmental Stochasticity)

Consider an annual plant species. Every year in the spring, a seed will either grow or remain dormant in the ground for another year. The decision of whether to grow or not is made before environmental conditions are known for that year. Conditions during growth can be "Good" or "Bad"; most years are Good. If it is a Good year, a germinating seed grows to a mature plant during the summer, produces many seeds, and then dies. If it is a Bad year, a germinating plant dies. Seeds that remain dormant are not affected by the type of year. A strategy for a plant specifies the proportion of its seeds that germinate each year.

We can take the performance measure to be number of seeds in the ground next winter that are descendants of a seed that is in the ground this winter. Consider first the performance of the strategy of having all seeds germinate. This strategy does very well in Good years, leaving many descendants, but leaves no descendants in Bad years. Since most years are Good and a mature plant produces many seeds, this strategy maximizes the mean number of descendant seeds left next year per current seed. However, any genotype that coded for this strategy would be wiped out in the first Bad year—so this strategy would not evolve. Instead, we expect a bet-hedging strategy to evolve: some seeds germinate, others remain dormant as insurance against Bad years (Cohen 1966). This strategy would be favored over immediate germination because it is more robust against seasonal variation.

General Discussion

In both Cases 1 and 3, there will be selection for robust strategies because of multiplicative effects. In Case 1, the probability of overwinter survival is the product of the probabilities of survival on each day. In Case 3, the number of descendants left far into the future is the product of the numbers left from year to year. In general, whether effects are multiplicative or additive depends both on the performance measure and on the spatial and temporal structure of the environment (McNamara et al. 2011).

At the other extreme, there can be selection to take risks. For example, in elephant seals a few males are able to monopolize most of the breeding females; thus, it may be worth it for a male to take big risks to become dominant.

From "Evolution and the Mechanisms of Decision Making," edited by Peter Hammerstein and Jeffrey R. Stevens. 2012. Strüngmann Forum Report, vol. 11, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01808-1.

In situations where the current habitat properties are unknown, there may be selection to be "optimistic" and take a risk that it is a good habitat (McNamara et al. 2011). In these cases there is no selection for robustness. In fact, adaptive strategies are far from robust.

In general more modeling work is needed on the question of whether it is adaptive to be robust. In the construction of models, it is important to broaden the worlds that are considered. It is impossible to build models of the real world. Nevertheless, more complexity is needed than has been used in most previous work. For example, consider the foraging strategy of a small bird that must gain enough energy to survive the night. If we model a world in which the bird "knows" the availability of food, and knows it will not change during the day, we get one prediction about risk-sensitive foraging. If we allow for uncertainty and change, we get another that is more robust (McNamara 1996).

Overfitting

Natural selection shapes the strategies of population members so that they are roughly adapted to the local environmental conditions experienced by ancestors. Mutation tends to work against this process, so we would expect the more recent past to have a greater influence than the more distant past on current population strategies. There is a tendency to "overfit" in that current individuals are adapted to the recent past and may not be adapted to future environmental conditions.

As an example, consider an environment that is stable apart from the occasional El Niño year. (El Niño is a warming of Pacific Ocean surface temperatures that occurs approximately every five years and influences global weather patterns.) We would expect a different distribution of strategies immediately after an El Niño event than after a run of normal years. After an El Niño event the population is more adapted to another such event, although it is likely to be a normal year. Even if the environment is stable over evolutionary time, the experience of an individual over its lifetime is liable to be different in detail to that experienced by any of its ancestors; so there will even be a tendency for short-term overfitting in evolutionarily stable environments.

In the El Niño example, robust mechanisms might be able to deal well with both normal years and El Niño years. But what are the opportunity costs of robust decision mechanisms? In periods of more stable or predictable environments, individuals using a robust mechanism would show a gradual proportional decrease in a population because there is no selection pressure for the robustness. Therefore, the timing of unpredictable events and the difference between the performance of other mechanisms compared to robust mechanisms is critical.

In terms of moment-to-moment decision making, a strategy specifies the rule by which an individual integrates its past experience in deciding how to respond to current stimuli. In a complex world, past experience is liable to be

From "Evolution and the Mechanisms of Decision Making," edited by Peter Hammerstein and Jeffrey R. Stevens.
2012. Strüngmann Forum Report, vol. 11, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01808-1.

unique in detail. Thus, the decision-making rule is dealing with a situation never experienced by any of its ancestors. How well it copes with this situation depends on the robustness of its rule. Overfitting will tend to produce rules that are not robust. This is analogous to statistical overfitting of data, in which over-parameterized models fit noise rather than the underlying pattern (cf. Goldstein and Gigerenzer 2009). Whether overfitting occurs may depend on the statistical properties of the environment, the learning capacities of the agent or organism, and the classes of rules on which selection acts. To the extent overfitting is likely to result in fragility (non-robustness), more general processes will be adaptive and likely selected, if they are able to evolve before catastrophe strikes.

Psychological Mechanisms and Robustness

Optimality modeling can tell us about broad selection pressures acting on the robustness of behavioral strategies, but to understand the fine detail we need to consider underlying psychological mechanisms (McNamara and Houston 2009).

There are both external and internal factors associated with robustness. For the former, from a computational or functional perspective, it can be useful to consider separating the chance of substantially negative outliers from the opportunities associated with variable, but typically beneficent, environments. Learning ontogenetically in the face of looming catastrophe would be highly maladaptive; therefore we have hard-wired, evolutionarily programmed mechanisms that tell us what to do. These can also be considered to be heuristics, rules of thumb (Hutchinson and Gigerenzer 2005), or modules (Kurzban, this volume). It is not clear how mutable they are in the light of experience (e.g., can we learn to climb some particular tree to flee a bear) or how generic policies (approach/withdrawal) interact with specific ones. We can use notions from control theory as ways of formalizing the effects of outliers (Doyle et al. 1989; Wald 1945). Finding ways to study these modules and getting insights into their historical appropriateness is an important task, lest this critical part of the architecture be just arbitrary. One route might be to compare defense mechanisms across species (Bolles 1970).

Some have distinguished at least two mechanisms of the internal environment that pertain to robustness: model-based and model-free control mechanisms (see Dayan, this volume). These controllers have different abilities to work in the face of computational and statistical uncertainty. Each embodies prior information in different ways (model-based control in a much richer manner than model-free control) and so can adapt to environments differently. A challenge is that the controllers interact richly, making it hard to tease apart their individual contributions.

The Cost-Benefit Trade-offs of Robustness

From an engineering standpoint, building robustness with respect to one dimension generally carries some sort of cost elsewhere for the system. For example, robustness to damage can be gained from adding redundant backup systems that mirror the functionality of primary systems. This approach is commonly used when primary system failure can lead to very costly outcomes, as in the case of hospital power generation systems and expensive and sophisticated machines, such as the space shuttle.

Thus, robustness generally entails a cost-benefit trade-off: the cost carried by the means used to gain robustness trades off against the benefit of the robustness gain. In the case of redundant systems, the benefit is gaining robustness to failure of the primary system, but at the additional cost of the backup system. Similarly, robustness to different dimensions of environmental perturbation trades off against one another. For instance, in designing an airplane, one might choose to include a redundant hydraulic system so that if the primary system is disabled (e.g., by a bullet or by a material fatigue-induced structural failure), the backup system can be relied upon to maintain the airplane's function. This increases survivability, which is a benefit. However, this benefit is achieved at the cost of maneuverability, due to the weight increase that results from the inclusion of the extra equipment.

Similarly, achieving robustness in biological systems generally carries some cost, often in the allocation of energy for the tissue required to increase robustness. The human ability to recognize faces is robust with respect to things such as viewing angle, lighting conditions, and so on. This, presumably, requires expensive nervous tissue to implement this functionality (for a discussion of the costs and benefits of larger brains, see Chittka and Niven 2009). Adding robustness, in this and other senses, can be understood as imposing some cost that must be made up either by sacrificing energy in some other domain or with additional collection of energy.

Costs can come in any number of forms. Consider a system that samples the environment in order to estimate some parameter, such as the mean of the distribution. Sampling from a large number of instances reduces the error (i.e., increasing the robustness of the estimate). However, reducing the error this way increases the time and energy spent sampling (Kareev, this volume).

The trade-offs inherent in achieving robustness help to explain why systems cannot be universally robust across all possible dimensions. Though there might be other reasons (e.g., phylogenetic constraints, local minima), selection will generally favor robustness mechanisms for which marginal robustness benefits outweigh their marginal costs, whatever those costs must be. Because many different robustness trade-offs must be addressed by any given design, selection can be understood to resolve a range of robustness trade-offs.

Note that robustness cannot be universally optimized because selection pressures change and there is uncertainty regarding the future environments that organisms face. However, the same uncertainties that make optimization impossible can and do result in both artificial and natural forms of "tolerance buffering," which serves to increase the robustness of the system to variation. For example, when designing and constructing a bridge, engineers build in a degree of safety margin assuming certain standard and extreme stresses. This is an intentional buffering for robustness that is possible in domains characterized by decision makers that are goal-directed, creative, and purposeful. Natural selection is characterized by fitness to the environment rather than creative intention, but we also see tolerance buffering in naturally selected biological systems. The liver, for instance, has substantially greater capacity than necessary for maintenance as well as the ability to regenerate or replace lost tissue. Robustness through excess capacity (for normal maintenance) is likely to prevail throughout biological design. While tuned to past environments in evolution and ontogeny, the capacity will often and at most times be in excess of that typically required to handle less common stress. There is an evolutionary advantage for mechanisms that buffer the system against higher levels of environmental variability, independent of whether the system is a natural or artificial one.

Robustness in Multi-Agent and Group-Level Decision Processes

The social environment offers a unique set of issues for robust decision making. Uncertainty associated with interaction with social agents may require robust mechanisms to cope with this variation. The strategic components of social interactions provide a particularly challenging environment for robustness because mechanisms must be robust not to an independent environment but one that responds directly to the organism.

In addition to providing obstacles to robustness, social agents can also enhance robustness by using group decision making. Group living enables animals to address many problems that would be difficult or impossible for single individuals. For example, individuals in groups can catch bigger prey or better protect themselves against predators (Krause and Ruxton 2002). Groups can afford forms of robustness that are not available to individuals. They permit things like learning from observation, which is another way of alleviating the requirement for individual acquisition (Oztop et al. 2006; see also Hammerstein and Boyd, this volume), herding, and also group rather than individual exploration of parameter spaces.

The recent introduction of self-organization theory into the behavioral sciences has led to many case studies which show that group living can also facilitate better decision making. The solution of cognitive problems that go beyond the capability of single animals is known as *collective intelligence* or *swarm intelligence* (Krause et al. 2010). A process definition is that two or more individuals independently, or at least partially independently, acquire information,

From "Evolution and the Mechanisms of Decision Making," edited by Peter Hammerstein and Jeffrey R. Stevens.
2012. Strüngmann Forum Report, vol. 11, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01808-1.

and these different packages of information are combined and processed through social interaction, which provides a solution to a cognitive problem.

There are many examples of swarm intelligence in invertebrates (particularly in social insects) and more recently in vertebrates. The perception in most case studies (particularly in invertebrates) has been that the individual animal is cognitively relatively simple and restricted in what it can achieve, whereas the group collectively is capable of solving difficult problems.

One of the first scientists to point out that individuals can benefit from collective decision making was the French mathematician and political scientist Nicolas de Condorcet. He assumed that each individual can either be right or wrong with a certain probability p. Provided that $p(right) > 0.50$, the probability of a correct collective decision of the group will increase as a function of group size, provided that the individuals that have the correct information are in the majority in the population.

Subsequent research has demonstrated, quite counterintuitively, that swarm intelligence does not require a majority of individuals who know the correct answer or, in fact, any individual to know the correct answer. This was first shown by Galton (1907) in an empirical study on humans based on what is now known as the *many-wrongs principle* (Bergmann and Donner 1964). The many-wrongs principle is often mentioned in the context of navigational problems where navigational accuracy is predicted to increase as a function of group size (Simons 2004). The assumption underlying the many-wrongs principle is that all individuals have a common target destination, but that each individual navigates with some error. If group members average over their directional preferences (through social interaction), then the error with which the group moves toward the target decreases as a nonlinear function of group size. In the example by Simons (2004), individual errors followed a normal distribution. However, the principle of the many wrongs producing a good overall decision is not restricted to a particular type of distribution. As long as the mean of the individual vectors approximates the target direction, there are different types of distribution that could produce a similar outcome (i.e., reducing navigational error with group size). Thus, group decision making provides a robust response to individual errors in the estimates.

Collective decision making is usually robust to changes in the environment, outliers, and loss of group members, but it also has costs and is based on a number of prerequisites, one of which is independence of individuals. For instance, if all individuals in a group have the same bias, then no degree of redundancy (i.e., no increase in group size) is likely to make the collective decision any better. In contrast, the greater the independence of individuals in a system, the greater the probability will be that (a) a solution will emerge and (b) decision-making quality will increase with group size.

The self-organized nature of collective behavior is a form of redundancy that protects against failure of system components. Self-organization means that the decision-making process is decentralized and therefore less vulnerable to localized damage than central-control systems, where the loss of only a few individuals can have a strong adverse effect if they are the leaders or central decision makers.

Robust Group Decision Making in Animals

In honeybees, *Apis melifera*, workers perform different tasks which correlate with their age and development. Younger bees usually work inside the hive building cells and feeding larvae, whereas older workers collect nectar and pollen outside the hive. The colony functions best when there is a certain ratio of bees to carry out the indoor and outdoor services. If, however, a large number of bees get killed while performing outdoor duties, then the colony can respond adaptively with younger workers developing more rapidly to take on outdoor tasks (Robinson 1992; Schulz et al. 1998). Likewise, if for some reason the number of young bees is drastically reduced, then older bees take up nest building tasks again. This process of worker allocation is not centrally controlled. It is achieved through the contact frequencies between the workers of different ages and task groups. During each contact between two workers, different hormones are exchanged; these hormones control worker development and thereby regulate the ratio of indoor and outdoor workers (Huang et al. 1998). This example clearly shows how a self-organized process can allow a robust response to the loss of workers, not only through redundancy but also through reorganization of task allocation.

Many examples of swarm intelligence come from the social insect literature, where quorum sensing has been identified as an important mechanism for decision making (some cases of quorum decision have been identified in vertebrates as well; Ward et al. 2008). Insect colonies sometimes need to find a new nest location because their old nest has either become too small or has been damaged. The problem that the colony needs to solve frequently takes the form of a complex trade-off between speed and accuracy of decisions. A proportion of scouts (individuals that explore the surroundings for suitable nest locations) leave the nest and, if successful, these individuals then try to recruit others to the new potential nest location they encountered. After the number of individuals in support of a particular nest location reaches a threshold (i.e., a quorum), the entire colony will favor this location (Franks et al. 2009). If speed is important, then the quorum threshold can be low, resulting in lower accuracy and reducing the probability of deciding in favor of the best new location among the available options. If speed is not a constraint, then the quorum threshold can be high and accuracy is increased (Franks et al. 2009). A similar type of quorum-based decision making has also been described in fish shoals that have to make a decision about which path to take to avoid danger or which leader to follow (Ward et al. 2008).

From "Evolution and the Mechanisms of Decision Making," edited by Peter Hammerstein and Jeffrey R. Stevens. 2012. Strüngmann Forum Report, vol. 11, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01808-1.

Group Decisions and Robustness in Science and Technology

One of the richest areas for research and application in group decision making is in science and technology. These communities have sophisticated methodologies for combining inputs from different sources to construct solutions or answers that are more robust, more reliable, and less prone to error. When software is sent out for beta-testing, a substantial population of unpaid testers explore it under diverse conditions, reflecting their diverse interests, and report failures that are used to produce corrections incorporated in the released version. This could be seen as a richer and more sensitive extension of the methodologies discussed above for decision making in groups.

In scientific investigations, it is standard to assume that a measurement or result derived using two or more different modes of instrumentation is more reliable, and that cross-checking in this way is commonly expected and accepted as a test for the "reality" or "non-artifactuality" of the measurement, result, or detected property or entity (Soler et al. 2011; Wimsatt 1981). Levins (1966) talks about "robust theorems"—results derivable in diverse models of a phenomenon using different assumptions, which thereby do not depend upon the details of the various specific models. Campbell (1958, 1966) talks about the importance of "triangulation" using different methods, and Campbell and Fiske (1959) expound the use of a "multi-trait, multi-method matrix" to correct for biases in methods of measurement or bad choice of indices for a trait. Physicist Richard Feynman (1967) contrasts "Babylonian" vs. "Euclidean" formal methodologies and argues the advantages of the multiply connected inferences in the former over the minimalist serial inferences of the latter. These all result in methods of scientific inquiry that provide robustness to the vagaries of individual biases in model creation and data collection.

In the larger social structure of science, demands for repeatability, public disclosure of methods, and the peer review system are all attempts to secure greater reliability through robustness. Are these methods foolproof? Of course not. If the various methods fail to be independent in relevant respects, this can compromise results. Wade (1978) reviewed twelve different models of group selection by various authors, representing advocates and opponents of group selection, including some of the most distinguished evolutionary biologists. His study demonstrates that their near-universal claims to show (robustly) that group selection is not a significant force is undercut by the fact that the models shared five simplifying assumptions that were (a) false and (b) each biased the case against group selection. Wimsatt (1980) analyzed these cases and found that the biased assumptions are products of simplifications resulting from heuristics used in reductionistic methods of formulating and solving problems, paralleling earlier work by Tversky and Kahneman (1974) on heuristics and biases in probabilistic inference. Any purported case of robustness is corrigible by demonstrations that the methods are not independent, or as Levins (1966) would say, "not a representative sample from the space of possible models."

From "Evolution and the Mechanisms of Decision Making," edited by Peter Hammerstein and Jeffrey R. Stevens. 2012. Strüngmann Forum Report, vol. 11, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01808-1.

This domain suggests that case studies of scientific inference, technology testing, the organization of laboratories, and the composition of peer review panels all provide rich possible sources for understanding methodologies for generating and testing robustness in group decision making. We would like to know particularly how such methods may fail, and how induced biases might be detected and corrected.

Concluding Remarks

Robustness is an integral concept in the evolution of biological organisms and nonbiological systems, and a principal driver in the evolution of different decision-making and control mechanisms. To understand the evolutionary origins of decision mechanisms, therefore, we must address the nature of robustness, the nature of environments, and the costs and potential benefits associated with robustness in these environments.

In this chapter we have highlighted a number of issues that are central to robustness. For example, we have argued that in organisms, redundancy in the mechanisms of decision making is often not achieved by duplicating mechanisms but rather through the cooperation and competition of somewhat different mechanisms. Our primary emphasis has been on the extent and type of variation that individual organisms or technologies face. This variation provides a source of perturbation and is therefore central to whether we expect natural selection to produce optimal strategies, flexible strategies, and/or robust strategies. Selection acts on the underlying psychological mechanisms which must implement the behavioral strategy, and different mechanisms are robust to different sorts of environmental perturbation and may have different costs. We have highlighted how robustness can be achieved via multi-agent decision making, in the hopes that this discussion will stimulate further study into these important issues.

From “Evolution and the Mechanisms of Decision Making,” edited by Peter Hammerstein and Jeffrey R. Stevens.
2012. Strüngmann Forum Report, vol. 11, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01808-1.