



Decision Theory: What "Should" the Nervous System Do?

Konrad Körding, *et al.*
Science **318**, 606 (2007);
DOI: 10.1126/science.1142998

The following resources related to this article are available online at www.sciencemag.org (this information is current as of March 19, 2008):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/318/5850/606>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/318/5850/606#related-content>

This article **cites 33 articles**, 6 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/318/5850/606#otherarticles>

This article appears in the following **subject collections**:

Neuroscience

<http://www.sciencemag.org/cgi/collection/neuroscience>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

12. M. L. Platt, P. W. Glimcher, *Nature* **400**, 233 (1999).
13. W. Schultz, *Annu. Rev. Psychol.* **57**, 87 (2006).
14. J. I. Gold, M. N. Shadlen, *Trends Cognit. Sci.* **5**, 10 (2001).
15. A. R. Damasio, *Sci. Am.* **271**, 144 (1994).
16. A. Bechara, H. Damasio, D. Tranel, A. R. Damasio, *Science* **275**, 1293 (1997).
17. A. Bechara, H. Damasio, A. R. Damasio, *Cereb. Cortex* **10**, 295 (2000).
18. A. D. (Bud) Craig, *Trends Neurosci.* **26**, 303 (2003).
19. N. D. Daw, J. P. O'Doherty, P. Dayan, B. Seymour, R. J. Dolan, *Nature* **441**, 876 (2006).
20. B. De Martino, D. Kumaran, B. Seymour, R. J. Dolan, *Science* **313**, 684 (2006).
21. G. F. Koob, M. Le Moal, *Science* **278**, 52 (1997).
22. M. Bateson, *Proc. Nutr. Soc.* **61**, 509 (2002).
23. C. J. Pietras, T. D. Hackenber, *J. Exp. Anal. Behav.* **76**, 1 (2001).
24. G. Dom, B. Sabbe, W. Hulstijn, B. W. van den Brink, *Br. J. Psychiatry* **187**, 209 (2005).
25. G. J. Madden, W. K. Bickel, E. A. Jacobs, *Exp. Clin. Psychopharmacol.* **7**, 284 (1999).
26. S. F. Coffey, G. D. Gudleski, M. E. Saladin, K. T. Brady, *Exp. Clin. Psychopharmacol.* **11**, 18 (2003).
27. N. M. Petry, T. Casarella, *Drug Alcohol Depend.* **56**, 25 (1999).
28. N. M. Petry, *Psychopharmacology (Berlin)* **162**, 425 (2002).
29. G. Ainslie, in *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person* (Cambridge Univ. Press, New York, 1992), pp. 56–95.
30. K. N. Kirby, N. M. Petry, W. K. Bickel, *J. Exp. Psychol. Gen.* **128**, 78 (1999).
31. S. H. Kollins, *Addict. Behav.* **28**, 1167 (2003).
32. S. Grant, C. Contoreggi, E. D. London, *Neuropsychologia* **38**, 1180 (2000).
33. R. Gonzalez, A. Bechara, E. M. Martin, *J. Clin. Exp. Neuropsychol.* **29**, 155 (2007).
34. A. Verdejo-Garcia, R. Vilar-Lopez, M. Perez-Garcia, K. Podell, E. Goldberg, *J. Int. Neuropsychol. Soc.* **12**, 90 (2006).
35. A. Verdejo-Garcia, C. Rivas-Perez, R. Vilar-Lopez, M. Perez-Garcia, *Drug Alcohol Depend.* **86**, 139 (2007).
36. B. B. Quednow et al., *Psychopharmacology (Berlin)* **189**, 517 (2007).
37. G. Dom, B. De Wilde, W. Hulstijn, B. W. van den Brink, B. Sabbe, *Alcohol. Clin. Exp. Res.* **30**, 1670 (2006).
38. R. Pirastu et al., *Drug Alcohol Depend.* **83**, 163 (2006).
39. A. Bechara et al., *Neuropsychologia* **39**, 376 (2001).
40. R. D. Rogers et al., *Neuropsychopharmacology* **20**, 322 (1999).
41. M. P. Paulus, N. Hozack, L. Frank, G. G. Brown, M. A. Schuckit, *Biol. Psychiatry* **53**, 65 (2003).
42. J. Monterosso, R. Ehrman, K. L. Napier, C. P. O'Brien, A. R. Childress, *Addiction* **96**, 1825 (2001).
43. J. R. Monterosso et al., *Hum. Brain Mapp.* **28**, 383 (2007).
44. M. P. Paulus et al., *Neuropsychopharmacology* **26**, 53 (2002).
45. K. I. Bolla et al., *Neuroimage* **19**, 1085 (2003).
46. A. Bechara, *Nat. Neurosci.* **8**, 1458 (2005).
47. J. McCartney, *Subst. Use Misuse* **32**, 2061 (1997).
48. D. R. Cherek, S. D. Lane, *Psychopharmacology (Berlin)* **157**, 221 (2001).
49. D. S. Leland, M. P. Paulus, *Drug Alcohol Depend.* **78**, 83 (2005).
50. D. S. Leland, E. Arce, J. S. Feinstein, M. P. Paulus, *Neuroimage* **33**, 725 (2006).
51. S. Rahman et al., *Neuropsychopharmacology* **31**, 651 (2006).
52. C. N. Ortner, T. K. MacDonald, M. C. Olmstead, *Alcohol Alcohol.* **38**, 151 (2003).
53. J. B. Richards, L. Zhang, S. H. Mitchell, H. de Wit, *J. Exp. Anal. Behav.* **71**, 121 (1999).
54. H. de Wit, J. L. Enggasser, J. B. Richards, *Neuropsychopharmacology* **27**, 813 (2002).
55. F. X. Vollenweider, M. E. Liechti, M. P. Paulus, *J. Psychopharmacol.* **19**, 366 (2005).
56. B. Reynolds, J. B. Richards, M. Dassinger, H. de Wit, *Pharmacol. Biochem. Behav.* **79**, 17 (2004).
57. J. McDonald, L. Schleifer, J. B. Richards, H. de Wit, *Neuropsychopharmacology* **28**, 1356 (2003).
58. S. D. Lane, D. R. Cherek, O. V. Tcheremissine, L. M. Liewing, C. J. Pietras, *Neuropsychopharmacology* **30**, 800 (2005).
59. S. D. Lane, O. V. Tcheremissine, L. M. Liewing, S. Nouvion, D. R. Cherek, *Psychopharmacology (Berlin)* **181**, 364 (2005).
60. J. Epstein et al., *Am. J. Psychiatry* **163**, 1784 (2006).
61. E. E. Forbes et al., *J. Child Psychol. Psychiatry* **47**, 1031 (2006).
62. A. Must et al., *J. Affect. Disord.* **90**, 209 (2006).
63. T. Christodoulou, M. Lewis, G. B. Ploubidis, S. Frangou, *Eur. Psychiatry* **21**, 270 (2006).
64. F. C. Murphy et al., *Psychol. Med.* **31**, 679 (2001).
65. A. Minassian, M. P. Paulus, W. Perry, *J. Affect. Disord.* **82**, 203 (2004).
66. R. Ladouceur, P. Gosselin, M. J. Dugas, *Behav. Res. Ther.* **38**, 933 (2000).
67. I. Blanchette, A. Richards, *J. Exp. Psychol. Gen.* **132**, 294 (2003).
68. M. P. Paulus, J. S. Feinstein, A. Simmons, M. B. Stein, *Biol. Psychiatry* **55**, 1179 (2004).
69. A. L. Krain et al., *J. Child Psychol. Psychiatry* **47**, 1023 (2006).
70. S. Reiss, R. A. Peterson, D. M. Gursky, R. J. McNally, *Behav. Res. Ther.* **24**, 1 (1986).
71. M. P. Paulus, M. B. Stein, *Biol. Psychiatry* **60**, 383 (2006).
72. A. Simmons, S. C. Matthews, M. B. Stein, M. P. Paulus, *Neuroreport* **15**, 2261 (2004).
73. M. P. Paulus, C. Rogalsky, A. Simmons, J. S. Feinstein, M. B. Stein, *Neuroimage* **19**, 1439 (2003).
74. P. L. Remijnse et al., *Arch. Gen. Psychiatry* **63**, 1225 (2006).
75. N. S. Lawrence et al., *Neuropsychology* **20**, 409 (2006).
76. M. M. Nielen, D. J. Veltman, R. de Jong, G. Mulder, J. A. den Boer, *J. Affect. Disord.* **69**, 257 (2002).
77. K. D. Fitzgerald et al., *Biol. Psychiatry* **57**, 287 (2005).
78. P. S. Sachdev, G. S. Malhi, *Aust. N. Z. J. Psychiatry* **39**, 757 (2005).
79. O. H. Turnbull, C. E. Evans, K. Kemish, S. Park, C. H. Bowman, *Neuropsychology* **20**, 290 (2006).
80. S. B. Hutton et al., *Schizophr. Res.* **55**, 249 (2002).
81. M. P. Paulus, M. A. Geyer, D. L. Braff, *Am. J. Psychiatry* **153**, 714 (1996).
82. M. P. Paulus, M. A. Geyer, D. L. Braff, *Schizophr. Res.* **35**, 69 (1999).
83. M. P. Paulus, W. Perry, D. L. Braff, *Biol. Psychiatry* **46**, 662 (1999).
84. K. Ludewig, M. P. Paulus, F. X. Vollenweider, *Psychiatry Res.* **119**, 293 (2003).
85. M. P. Paulus, L. Frank, G. G. Brown, D. L. Braff, *Neuropsychopharmacology* **28**, 795 (2003).
86. S. Stroup et al., *Schizophr. Res.* **80**, 1 (2005).
87. A. D. Redish, *Science* **306**, 1944 (2004).
88. I would like to acknowledge the help of E. Arce, D. Leland, S. Matthews, M. Wittmann, and A. Simmons during the preparation of the manuscript. This research was supported by grants from the National Institute on Drug Abuse (R01DA016663 and R01DA018307) and a U.S. Department of Veterans Affairs merit grant.

10.1126/science.1142997

REVIEW

Decision Theory: What “Should” the Nervous System Do?

Konrad Körding

The purpose of our nervous system is to allow us to successfully interact with our environment. This normative idea is formalized by decision theory that defines which choices would be most beneficial. We live in an uncertain world, and each decision may have many possible outcomes; choosing the best decision is thus complicated. Bayesian decision theory formalizes these problems in the presence of uncertainty and often provides compact models that predict observed behavior. With its elegant formalization of the problems faced by the nervous system, it promises to become a major inspiration for studies in neuroscience.

Evolutionary psychology has found that many human behaviors can be well understood assuming adaptation of psychology to the past social environment of humans

[e.g., (1)]. Similarly, ethology, the study of animal behavior [e.g., (2)], has shown that many of the properties of the nervous system and the bodies of animals are remarkably well adapted to their eco-

logical niche. These disciplines have shown that, over the course of evolution, animals are often endowed with solutions to common problems that are close to optimal [(1), but see (3)]. Many studies in neuroscience analyze low-level processes. For example, researchers study how animals control their limbs, how they infer events in the world, and how they choose one of several possible rewards. Such processes may have remained conserved for very long periods of time. We can thus expect the solution used by the nervous system for such problems to be close to optimal.

Normative models formalize how the idea of adaptation predicts properties of the nervous system. These models assume that a process has

Department of Physical Medicine and Rehabilitation, Department of Physiology, and Department of Applied Mathematics, Institute of Neuroscience, Northwestern University and Rehabilitation Institute of Chicago, Room O-922, 345 East Superior Street, Chicago, IL 60611, USA. E-mail: kk@northwestern.edu

an objective (e.g., walk using least energy). Such an objective is typically formalized by a function (“utility”) that defines how good the solution is (e.g., the energy used). By combining the utility function with constraints (e.g., properties of muscles), it is possible to derive the best possible solution to the problem, usually using computer simulations. If the prediction of the model is matched by biology, it is concluded that, indeed, we have understood the purpose of a system. Normative models thus ask the “why?” question and formalize the ultimate purpose of a system. In this review, I focus on normative approaches to decision making; the nervous system often comes up with near-optimal decisions in a world characterized by uncertainty. To understand the nervous system, descriptive knowledge of how the nervous system works should be combined with normative knowledge of which problems it solves.

Decision Theory

The purpose of the central nervous system is to make decisions so that we can thrive by interacting successfully with our environment (Fig. 1A). When we play darts, we need to decide which position on the dartboard to aim at (Fig. 1B). The dartboard assigns a score to any possible dart position and thus defines the outcome. The objective of playing darts is to obtain as many points as possible. Within decision theory, such an objective is called a utility function [e.g., (4)]. A utility function $U(\text{outcome})$ measures how good or bad any possible decision outcome is. If dart players could choose where the dart will hit the board, they would choose the position that yields the most points and would thus maximize utility.

Although we can freely make decisions, we cannot directly choose the decision outcomes. If we always aim for the same position a , say the center of the bull’s eye, and throw many darts, we will produce a distribution of dart positions, x , on the dart board (Fig. 1C, inset). Within decision theory, this probability distribution is denoted $p(\text{outcome} = x | \text{decision} = a)$. If we aim at the position on the board that gives the highest score, we may instead hit a neighboring area of the dartboard and receive a low score. Depending on the position we aim at, different scores become more or less likely. This is a special case of a general problem in decision theory: Outcomes depend on decisions in a probabilistic way.

To derive the most beneficial decision, it is necessary to combine the utility function with knowledge of how our decisions affect potential outcomes. The expected utility is

$$E[\text{Utility}(\text{decision})] = \sum_{\text{possible outcomes}} p(\text{outcome} | \text{decision}) U(\text{outcome})$$

The best decision is then defined as the one that will maximize the expected utility (5).

The decision theoretic approach can be used whenever we know how decisions are related to outcomes and we know the utility of the outcomes. We can apply this framework to the example of darts playing. If we had low motor noise, we would be best off aiming at the triple 20 (Fig. 1C). In contrast, if we have a realistic value of motor noise, the best point at which to aim is not the point of maximal score but is to the lower left of the board (Fig. 1D, marked by dart). Both the behavior of advanced amateur players who have moderate motor variance and the behavior of professional players who have low motor variance are predicted by this decision theoretic approach. A range of recent studies of decision-making have analyzed situations that are analogous to playing darts (6). Such simple decisions are well predicted from the assumption that people solve their decision problems in a fashion that is close to optimal. The approach also applies to many animal behaviors: Animals need to choose to forage or rest, fight or flight, continue moving or freeze in place. Moreover, any behavior of an animal is, in some abstract way, a decision. The nervous system chooses one behavior from the set of all possible behaviors. Decision theory is thus a fundamental formalization of many problems that are solved by the nervous system and studied in neuroscience.

A number of recent studies, under the umbrella name of neuroeconomics, have started to analyze how the nervous system represents and computes with a utility function (7–9). Because

utility is central to any decision, it is important to understand how the nervous system represents reward. Utilities are important for the nervous system because it uses them to act successfully. It has been shown that the nervous system represents changes in expected utility in a way predicted by a major learning theory, reinforcement learning (10). Traditionally simple rewards, such as monetary or food incentives, are used in experiments that analyze how the nervous system represents utilities. However, insights from evolutionary biology (1) predict that many different factors will influence the utility of decisions. Indeed, it has been shown that the nervous system exhibits a reward signal when someone else who cheated in a game is punished (11). Future research will have to uncover the full complexity of how utility functions are represented and used by the nervous system.

Bayesian Statistics

Deciding seems easy: Choose the action that is associated with the highest expected utility. However, the probability of an outcome given the decision is difficult to estimate. To do so, we need to predict how the world will change until the outcome and how any decision would affect the world. This prediction can only be probabilistic because we have uncertainty about the properties of the world, stemming, for example, from noisy perception (12).

Bayesian statistics defines how uncertain pieces of information may be combined into a

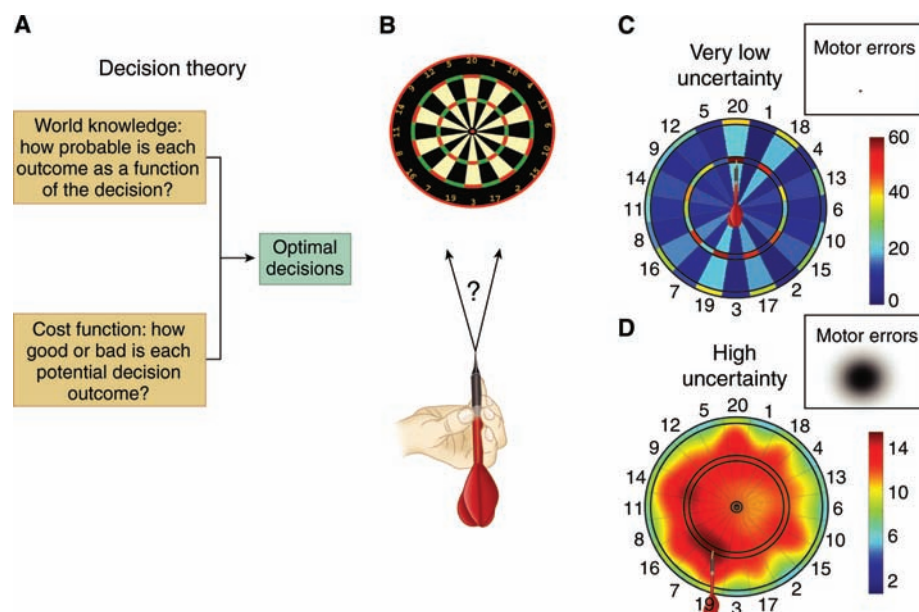


Fig. 1. Decision theory. (A) To make optimal decisions, we need to combine what we know about the world with our utility function measuring how good or bad potential outcomes are. (B) In the example of playing darts, we need to decide where to aim. (C) As a function of the aiming point, the expected score is shown for an unbelievably good darts player with almost no movement errors. (D) As in (C) but for a mediocre darts player with large motor errors.

Decision-Making

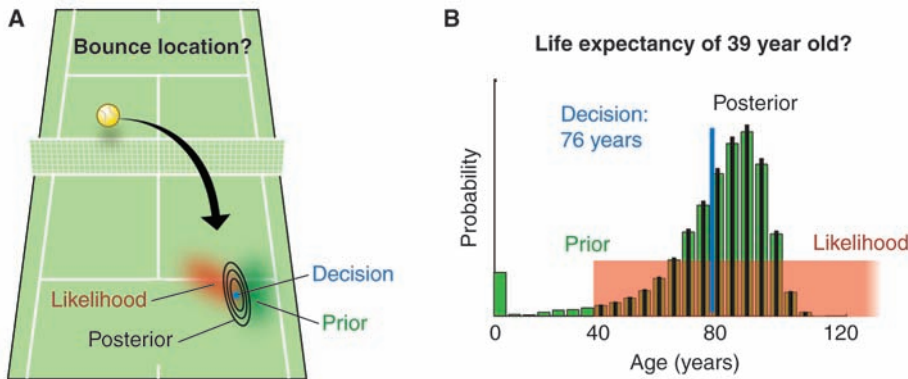


Fig. 2. (A) In the example of tennis, people need to combine what they know from before (prior, green) with what they currently see (likelihood, red). That way we can estimate the posterior (black contour lines) to make an optimal perceptual decision (blue). **(B)** Similarly if we estimate the life expectancy of a person who is 39 years old, we need to combine what we know from before (prior, histogram of lifetimes, green) with our new information (person survived 39 years, likelihood, red) to come up with an optimal estimate.

joint estimate. New information (called a likelihood) needs to be combined or integrated with information from the past (called a prior). Similar problems occur when information from several cues, for example, proprioceptive and visual, needs to be combined into a joint estimate. Bayesian decision theory (13), the use of Bayesian statistics in a decision framework, defines how our beliefs should be combined with our utility function. Because most if not all of our decisions are made in the presence of uncertainty,

understanding the way the nervous system deals with uncertainty is central to understanding its normal mode of operation.

Integration of priors and likelihoods. To calculate the probabilities of outcomes, it is often necessary to update our belief from the past (prior) with new knowledge (likelihood). For example, when we play tennis it is helpful to estimate where the ball will land. The visual system, although noisy, still provides us with an estimate or a likelihood of where the ball

will land (sketched in red in Fig. 2A). This knowledge may be combined with information obtained from experience; the positions where the ball may land are not uniformly distributed over the court. The locations may be clustered near the boundary lines, where it is most difficult to return the ball. This distribution of positions is called the prior (sketched in green in Fig. 2A). Bayes's rule states that how the probability of the ball landing at position x given our observation o (posterior) needs to be estimated as

$$\underbrace{p(x|o)}_{\text{posterior}} = \underbrace{p(x)}_{\text{prior}} \underbrace{p(o|x)}_{\text{likelihood}} / \underbrace{p(o)}_{\text{prior}}$$

Recent studies have analyzed such combinations in simple integration problems. Sensorimotor integration, force estimation, timing estimation, speed estimation, the interpretation of visual scenes, just to name a few, have been analyzed (14, 15). Together, these studies demonstrate that people intuitively combine prior knowledge with new evidence in a way predicted by Bayesian statistics.

Bayesian methods also apply to decision-making in cognitive contexts (16). What would be your guess of the life expectancy of a 39-year-old? People can use two sources of information to answer this question. They may use the prior, the distribution of lifetimes (Fig. 2B, green). They may also use the likelihood, this person must have survived the first 39 years of his or her life (Fig. 2B, red). With Bayes's rule, we can combine these two pieces of information, just as in the example of tennis (Fig. 2B, black). We thus estimate a life expectancy of about 76 years (using the negative square error as the utility function). Human participants exhibit cognitive behaviors that are close to optimal predictions. The same approach has been used to successfully predict human estimates for many other everyday cognition problems (16). People incorporate knowledge about the probability distributions into estimates in a fashion that is predicted by Bayesian decision theory.

Cue combination. Estimation will often depend on two different cues. For example, we may see and feel an object and use both senses to infer the properties of the object. Bayesian statistics allows us to solve these problems with the same mathematical framework used for the combination of prior and likelihood. A couple of recent studies have examined how subjects solve such cue combination problems. For example, the combinations of visual and auditory information and visual and tactile information, as well as within modality cue combination (e.g., texture and disparity), have been studied. In such cases, cues are combined in a fashion that is close to the optimum prescribed by Bayesian statistics (14, 15, 17, 18).

Although there is strong evidence that animals represent their degree of uncertainty and

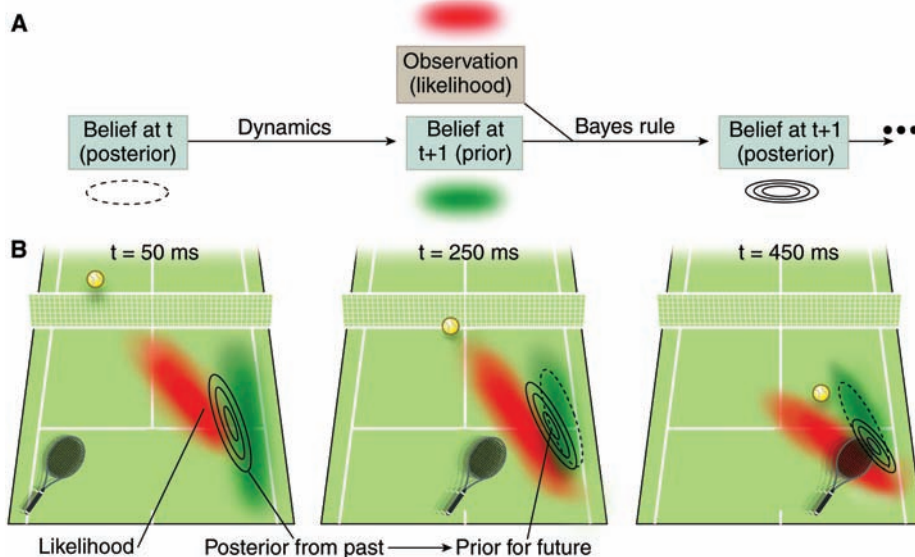


Fig. 3. Integration of information over time. (A) A diagram of a Kalman filter is shown. At any point of time t , the person has a belief about the state of the world. The person then updates this belief with a model of the dynamics of the world (e.g., gravity) to calculate the belief at the next point of time. This belief (prior) is then combined with new sensory information (likelihood) using Bayes's rule to calculate the belief at the next time step. The ellipses indicate probability distributions sketched in (B). **(B)** To estimate the position of a ball hitting the ground, people continuously update their beliefs with incoming sensory information, yielding precise estimates. The posterior of the previous time step is the prior for the new one: The dashed line indicating the previous posterior is identical to the one standard deviation line of the prior (green).

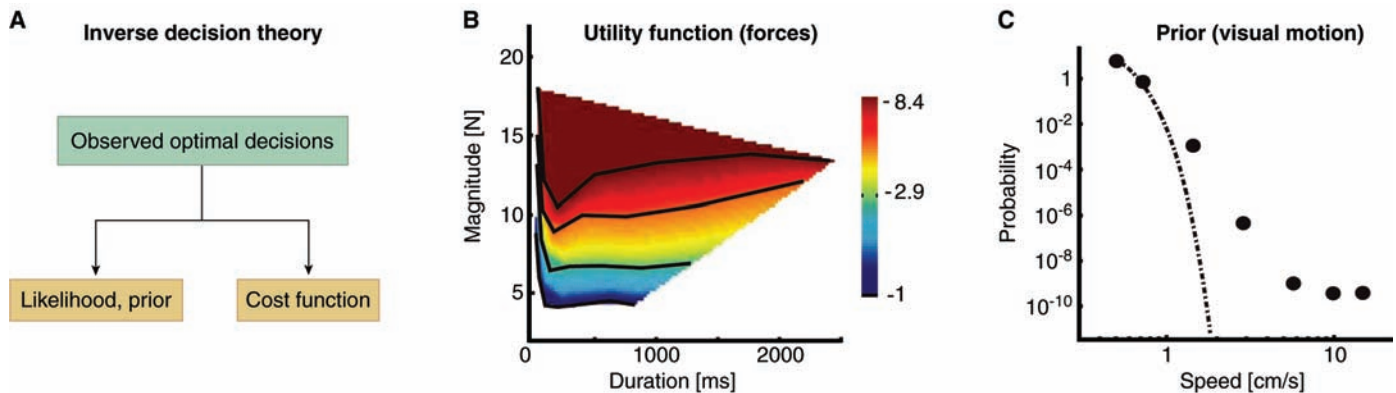


Fig. 4. Inverse decision theory. **(A)** In inverse decision theory, it is assumed that people behave optimally. From a sequence of optimal behavior, the priors, likelihoods, and utility functions people use are inferred through various computational methods. **(B)** In motor control, the utility function of producing forces of varying duration and magnitude have been calculated. **(C)** For visual decisions, people assume that small velocities are much more likely than large velocities.

combine different pieces of information into a joint estimate, there is still relatively little known about how the nervous system implements such computations. Recently, a couple of theoretical studies proposed how the nervous system may represent probabilities and combine them into estimates (19, 20). What is exciting about these studies is that they make clear testable predictions of how the nervous system may represent uncertainty. Understanding this representation is key for neuroscience because animals essentially make all their decisions in the presence of uncertainty.

Integration of information over time. The state of the world and our information about it is continually changing. In the example of tennis, we only used an estimate at a single point of time. In reality, tennis players continuously observe the ball's motion, updating their beliefs about the position where the ball will land in a continuous manner. The Kalman filter formalizes how such a process may work. The filter uses knowledge about the dynamics of the world to convert its belief about the state of the world at the previous instance of time into a belief at a future point in time. For example, in the tennis game we expect the ball to move because of its forward momentum and to accelerate because of gravity. The resulting belief (prior, Fig. 3, green) is then combined with new sensory information (likelihood, Fig. 3, red) to produce an updated belief (posterior, Fig. 3, ellipses). This way the Kalman filter updates its beliefs over time. In the case of playing tennis, such a strategy predicts that, as the ball flies, our estimate of the landing position will progressively be updated and usually become more precise. It is thus possible to make an efficient decision in situations that change continuously over time.

A range of recent studies have probed the strategies used by human participants in such situations. The way that people estimate the position of their hand in the dark is well predicted

by Kalman filtering (21). Similarly, when people balance a pole they also seem to use such strategies (22). Moreover, the way the nervous system represents changing muscle properties can be understood by assuming a Kalman filtering strategy (23, 24). These studies demonstrate that when people integrate information over time to make simple sensorimotor decisions, they seem to do so in a fashion that is consistent with the optimal Bayesian solution.

The nervous system constantly integrates information over time, and a range of new studies analyzed how it does so (25). In many such experiments, one of two stimuli is given, for example either a stimulus that moves to the right or a stimulus that moves to the left. If the stimulus is sufficiently noisy, the nervous system needs to integrate information over an extended period of time to make a good decision. Neurons were found that exhibit activities that correlate with the predicted process of optimal information integration over time. The nervous system takes into account probabilistic knowledge of potential rewards when integrating evidence for decision-making (26). The resulting models are particularly useful because they have a normative component (optimal integration of evidence) while having a straightforward descriptive component (neurons can integrate inputs over time).

Inverse Decision Theory

If people make decisions optimally, the mechanism of decision theory may be inverted: Computational techniques [e.g., (27)] are used to infer which priors, likelihoods, and utility functions the participants used to make their decisions. For example, a utility function with a few free parameters may be proposed, and the parameters may be fit to human behavior. Experimental economics (28) has extensively asked which utility functions people are optimizing. Only recently has the study of neuroscience and low-level decision-making started asking which

priors are used and which utility functions are optimized.

Many studies in neuroscience analyze motor control, an area where decision-making is key. For example, would you rather carry a 2-kg weight for 1 min or a 1-kg weight for 2 min? We intuitively make repeatable choices in such situations that are of relevance to everyday life. Such a utility function was recently inferred by using inverse decision theory (29) (Fig. 4A). This utility function is highly nonlinear in force magnitude and duration and is more complicated than previously proposed utility functions (30). This highlights a problem in decision theory: Frequently, good fits to behavior may be obtained with wrong utility functions. Inverse decision theory can thus be seen as a way of searching for violations of the assumptions made when building a decision theoretic model.

The framework of inverse decision theory also allows the analysis of which priors and which likelihoods are used by people. Studies indicate that people underestimate the speed of visual motion (31). This has been argued as the result of using a Gaussian prior for interpreting that low speeds are most likely. But why should people use a Gaussian distribution? We do not know which prior would be optimal in real life, although some recent progress in the statistics of natural scenes may lead that way (32). In a recent experiment, the prior used by human participants was measured by using inverse decision theory (33) (Fig. 4A). The prior is not Gaussian and is rather similar to an exponential function. This may inspire future experiments to characterize how the nervous system implements such a prior.

The strength of inverse decision theory in allowing for a wide range of possible utility functions is also its weakness. Inverse decision theory will always yield a utility function, likelihood, or prior for which the actually observed behavior is optimal. If the results differ from those assumed by a previous decision theoretic model, we can falsify this model. However, the

results should only be the basis of a new model if we can understand (and test) how the inferred functions derive from properties of the world. Similar problems also appear in other decision theoretic models that do not explicitly use inverse decision theory. The theoretician may fiddle with the decision theoretic model, trying different utilities, likelihoods, and priors, until there are good fits to human performance. As in all other models that explain data, overfitting is also a problem for decision theoretic models.

Inverse decision theory allows for the estimations of the used utility functions, priors, and likelihoods, which may alleviate the search for their neural representation. For example, if we know how utility functions depend on ideas of fairness when playing games, it is possible to search for brain areas that represent this aspect of the utility function (11). Searching for the neural representation of a utility function that has been proposed on theoretical grounds, but is irrelevant for human behavior, may miss important aspects of decision-making. When searching for the representation of priors or utility functions in the nervous system, it seems central to know the form of the priors and utility functions that are actually used by human participants.

Discussion

The world is complicated, and consequently so is deciding. Models in neuroscience typically analyze simple relationships between variables. However, at least our high-level behavior is characterized by structural relationships. Events in the world have causes, and we naturally interpret events in terms of cause and effect. A few studies of Bayesian statistics over the past couple of years have started to address the issue of how people may be able to infer the structure of the world (34, 35). Structure implies that not only features but their relationships play a fundamental role. This concept has long been at the heart of cognitive science. How people solve complicated real-world problems needs to be understood. A lot of recent progress in machine learning aims at inferring the structure of the world from real data (36), a process that people perform effortlessly. The study of decisions in neuroscience can draw upon advances in machine-learning to make interesting new predictions.

The decision theoretic approach may be limited in several ways. Humans should behave suboptimally for ethologically new kinds of decisions that are not repeated enough to allow for learning [but see (37)]. This may, for example, be relevant to the way people participate in lotteries (38). Using Bayes's rule to combine pieces

of information is the best mathematical solution to any information combination problem and thus always has been the best solution. Similarly, moving efficiently has always been beneficial to animals. The simple low-level properties of neural decision-making should thus be expected to be close to optimal.

Decision theory formalizes how animals should decide and thus does not directly make predictions of how the nervous system should implement the algorithm leading to such decisions (39). Countless different implementations may lead to the same optimal decision rules. However, the normative approach is not limited to decision theory. There may equally be costs and benefits to implementing algorithms in various ways. For example, having many long connections between neurons may use volume and energy and lead to slow information transmission. The idea of wiring length minimization explains well the placement of neurons in cortical maps (40) as well as within the body of the nematode *Caenorhabditis elegans* (41). Similarly, it may be argued that visual neurons should faithfully represent the world with the fewest spikes (least metabolic cost), an idea that predicts many properties of sensory representations (42). Normative ideas may even apply to cellular properties. The density of sodium channels in the squid giant axon may be understood from the idea that the squid giant axon should rapidly transmit action potentials used to allow the squid to flee from a predator (43, 44). Normative models and decision theory in particular offer ways of formalizing important problems that the nervous system needs to solve. Models in neuroscience should seek to explain the wealth of available experimental data and also incorporate knowledge of the problem solved by the system.

References

1. J. H. Barkow, L. Cosmides, J. Tooby, *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (Oxford Univ. Press, New York, 1992).
2. E. R. Weibel, C. R. Taylor, L. Bolis, *Principles of Animal Design: The Optimization and Symmorphosis Debate* (Cambridge Univ. Press, Cambridge, 1998).
3. S. J. Gould, R. Lewontin, *Proc. R. Soc. London Ser. B* **205**, 581 (1979).
4. D. Bernoulli, *Comentarii academiae scientiarum imperialis Petropolitanae* (for 1730 and 1731) **5**, 175 (1738).
5. J. v. Neumann, O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton Univ. Press, Princeton, NJ, 1944).
6. L. T. Maloney, J. Trommershäuser, M. S. Landy, in *Integrated Models of Cognitive Systems*, W. Gray, Ed. (Oxford Univ. Press, New York, 2006).
7. P. W. Glimcher, A. Rustichini, *Science* **306**, 447 (2004).
8. W. Schultz, *Curr. Opin. Neurobiol.* **14**, 139 (2004).
9. L. P. Sugrue, G. S. Corrado, W. T. Newsome, *Nat. Rev. Neurosci.* **6**, 363 (2005).
10. B. Seymour *et al.*, *Nature* **429**, 664 (2004).
11. D. J.-F. de Quervain *et al.*, *Science* **305**, 1254 (2004).
12. D. M. Green, J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
13. D. M. Wolpert, Z. Ghahramani, in *The Oxford Companion to the Mind*, R. L. Gregory, Ed., in press.
14. D. Knill, W. Richards, *Perception as Bayesian Inference* (Cambridge Univ. Press, New York, 1996).
15. K. P. Kording, D. M. Wolpert, *Trends Cogn. Sci.* **10**, 320 (2006).
16. T. L. Griffiths, J. B. Tenenbaum, *Psychol. Sci.* **17**, 767 (2006).
17. M. O. Ernst, H. H. Bulthoff, *Trends Cogn. Sci.* **8**, 162 (2004).
18. K. Kording *et al.*, *PLOS One* **2**, 10.1371/journal.pone.0000943 (2007).
19. W. J. Ma, J. M. Beck, P. E. Latham, A. Pouget, *Nat. Neurosci.* **9**, 1432 (2006).
20. A. Pouget, P. Dayan, R. S. Zemel, *Annu. Rev. Neurosci.* **26**, 381 (2003).
21. D. M. Wolpert, Z. Ghahramani, M. I. Jordan, *Science* **269**, 1880 (1995).
22. B. Mehta, S. Schaal, *J. Neurophysiol.* **88**, 942 (2002).
23. J. W. Krakauer, P. Mazzoni, A. Ghazizadeh, R. Ravindran, R. Shadmehr, *PLoS Biol.* **4**, e316 (2006).
24. K. P. Kording, J. B. Tenenbaum, R. Shadmehr, *Nat. Neurosci.* **10**, 779 (2007).
25. J. I. Gold, M. N. Shadlen, *Annu. Rev. Neurosci.* **30**, 535 (2007).
26. T. Yang, M. N. Shadlen, *Nature* **447**, 1075 (2007).
27. A. Y. Ng, S. Russell, in *Proceedings of the Seventeenth International Conference on Machine Learning* (Morgan Kaufmann, Burlington, MA, 2000), pp. 663–670.
28. J. H. Kagel, A. E. Roth, *The Handbook of Experimental Economics* (Princeton Univ. Press, Princeton, NJ, 1995).
29. K. P. Kording, I. Fukunaga, D. Wolpert, *PLoS Biol.* **2**, e330 (2004).
30. E. Todorov, *Nat. Neurosci.* **7**, 907 (2004).
31. Y. Weiss, E. P. Simoncelli, E. H. Adelson, *Nat. Neurosci.* **5**, 598 (2002).
32. B. Y. Betsch, W. Einhauser, K. P. Kording, P. Konig, *Biol. Cybern.* **90**, 41 (2004).
33. A. A. Stocker, E. P. Simoncelli, *Nat. Neurosci.* **9**, 578 (2006).
34. J. B. Tenenbaum, T. L. Griffiths, *Behav. Brain Sci.* **24**, 629 (2001).
35. J. B. Tenenbaum, T. L. Griffiths, C. Kemp, *Trends Cogn. Sci.* **10**, 309 (2006).
36. A. Yuille, D. Kersten, *Trends Cogn. Sci.* **10**, 301 (2006).
37. J. J. Koehler, *Behav. Brain Sci.* **19**, 1 (1996).
38. D. Kahneman, A. Tversky, *Econometrica* **47**, 313 (1979).
39. D. Marr, *Vision: A Computational Approach* (Freeman, San Francisco, 1982).
40. D. B. Chklovskii, A. A. Koulakov, *Annu. Rev. Neurosci.* **27**, 369 (2004).
41. B. L. Chen, D. H. Hall, D. B. Chklovskii, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4723 (2006).
42. B. A. Olshausen, D. J. Field, *Curr. Opin. Neurobiol.* **14**, 481 (2004).
43. A. L. Hodgkin, *Philos. Trans. R. Soc. London Ser. B* **270**, 297 (1975).
44. T. D. Sangrey, W. O. Friesen, W. B. Levy, *J. Neurophysiol.* **91**, 2541 (2004).

10.1126/science.1142998