
The Biology of Cooperative Decision-Making: Neurobiology to International Relations

5

Nicholas D. Wright

Introduction

To negotiate, cooperate or compete successfully with another, we should know what motivates them and how they make decisions. Neuroscience combined with psychology and economics tells us much about both this human motivation and decision-making. In this chapter, I describe three aspects of this neuroscientifically grounded account of decision-making that are central to negotiation and describe how each impacts on international negotiation and cooperation amongst states. Of course, neuroscience is no panacea, but we need the best evidence to negotiate, and neuroscience provides an important extra source of evidence.

In this chapter, I first discuss the broader biologically informed understanding of decision-making that draws on neuroscience, biology, psychology and economics [called neuroeconomics by some authors (Glimcher and Rustichini 2004; Glimcher and Fehr 2013)] and why it's arisen now. Second, I examine evidence from biology and neuroscience about how human cooperation emerges and is controlled. Third, I examine the neural bases of the fairness motivation and their importance in international negotiation. Fourth, I describe the neural phenomenon

of “prediction error” that affects the impact of our actions on others and how they will decide to respond to our actions. Fifth, I take a step back to give four simple rules for using this understanding of individual human decision-making to address policy issues in international negotiation. I give historical cases and practical policy recommendations throughout.

Combining Economics, Psychology and Neuroscience to Understand Decision-Making

Accounts of choice based in rational choice theory (RCT) (von Neumann and Morgenstern 1944) have dominated much of economics since the mid-twentieth century and more recently much of political science. The core concept in RCT is that an agent's choices are consistent, which is what makes the agent “rational”. RCT models individual choices through accounts such as expected utility theory, and models social choices through game theory. But although providing some useful tools, RCT fails to predict many aspects of human choice. To improve these models, over the past three decades, a subfield of economics, called behavioural economics, has aimed to “increase the explanatory power of economics by providing it with more realistic psychological foundations” (Camerer and Loewenstein 2004). However, “it is important to emphasize that the behavioural economics

N.D. Wright, M.R.C.P., Ph.D. (✉)
Carnegie Endowment for International Peace, 1779
Massachusetts Avenue, Washington, DC 20036, USA

approach extends rational choice and equilibrium models; it does not advocate abandoning these models entirely” (Ho et al. 2006). This combination of economics and psychology has, for example, sought to modify expected utility theory with prospect theory (Kahneman and Tversky 1979) and game theory with behavioural game theory (Camerer 2003)—but many core aspects of decision-making are still not captured.

Biologically based, neuroscientific approaches to choice have a long theoretical and empirical tradition, for instance, the vast literature on associative learning (Thorndike 1911; Mackintosh 1983). Over the past decade or so, this has been added to the combination of economics and psychology—to provide an extra source of evidence to understand decision-making (Glimcher 2004; Glimcher and Rustichini 2004; Camerer et al. 2005).¹ In this new field, the main object of interest is the study of value-based decision-making, that is, when an agent chooses from several alternatives based on the subjective values it places upon them. This interdisciplinary approach permits the introduction of new richness and robustness into models of human behaviour, within a mathematically specifiable and empirically grounded framework.²

Why has this arisen now? The advances in our understanding of human decision-making over the past decade were made possible by new, non-invasive brain imaging technologies. The key new technology has been functional magnetic resonance imaging (fMRI). fMRI measures changes in brain activity, through tightly coupled changes in local blood flow (Frackowiak et al. 2004), whilst individuals actually make decisions. The reason that these new technologies have precipitated such rapid advances in our understanding of decision-making is the neural scale on which they work—they provide data on

the level of systems within the brain, enabling us to link the vast existing neuroscientific literature from animals and humans directly to human behaviours previously described by psychology and economics. This neuroscientific grounding in particular helps us choose between competing explanations at the behavioural level (O’Doherty et al. 2007), it provides an additional independent source of evidence that increases the robustness of the conclusions (Wilson 1999) and it enhances our prior belief about the generalisability of findings across cultures that is crucially important in international negotiation. I address these and other general issues further in the Discussion.

Cooperation

A classic game capturing the tension between cooperation³ and self-interest is the prisoner’s dilemma game (PDG) (Flood and Drescher 1950). Consider two prisoners brought in for questioning by the KGB and placed in separate cells. If both stay silent (i.e. cooperate), they both receive 1 year in prison. If they both accuse the other (i.e. defect) they both get 4 years in prison. If one stays silent and the other defects, the cooperator gets 10 years in prison and the defector gets off scot-free. Game theory specifies that defection is the only rational choice, because it is superior whatever the other’s choice. However, if instead of both defecting the two players could cooperate, then they would receive a mutually more beneficial outcome.

Against the expectation of game theory, humans in the laboratory cooperate about half the time, even in a one-shot anonymous PDG (Kagel and Roth 1995; Camerer 2003). Of course, self-interest is also a motivation: individuals also respond to incentives, for instance raising the tempting payoff to defect (Kagel and Roth 1995; Camerer 2003). Humans are driven by both

¹ This may be referred to as neuroeconomics.

² Before continuing, I do not want to give the impression that RCT has little descriptive power in all games (e.g. the matching pennies game where individuals must keep their opponents guessing) (Camerer 2003). Furthermore, even where RCT does not well predict behaviour, it can give a useful conceptual perspective and mathematical framework.

³ Here I define cooperation as the voluntary acting together of two or more individuals that brings about, or potentially brings about, ends that benefit one, both or all, which are over and above the benefits arising from individualistic behaviour (Dugatkin 1997; Brosnan and de Waal 2002).

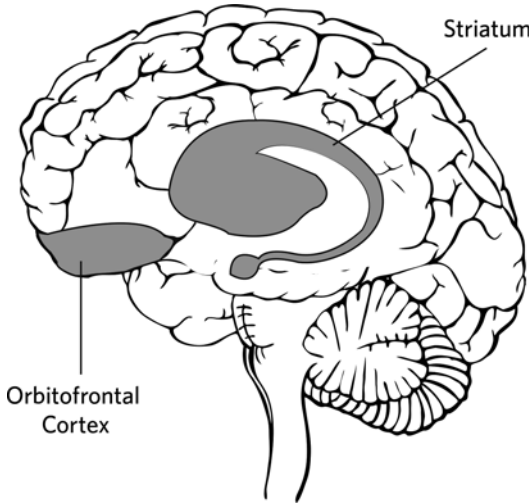


Fig. 5.1 Key brain regions for reward-based decision-making are the striatum and orbitofrontal cortex (OFC). Cooperation engages reward mechanisms in the brain

cooperation and self-interest—and both are based in their biology. This presents a different account of human motivation to that in RCT.

Neural Bases of Cooperative Behaviour

Humans and other animals have sophisticated neural machinery for reward-based decision-making, for example, to gain juice (in animals and humans) or money (in humans), in which it is well established that two key brain regions are the striatum and the orbitofrontal cortex (OFC) (O’Doherty 2004; Glimcher and Fehr 2013). As discussed below, these same brain structures are also implicated in the human drive to cooperate (Fig. 5.1).

One can study people in the brain scanner whilst they play the PDG or similar games for money. This shows that reward-related activity in the ventral striatum and OFC is elicited by mutual cooperation in an iterated PDG (Rilling et al. 2002) and in the closely related trust game (King-Casas et al. 2005). In the “trust game”, one player is given an amount of money (e.g. \$20) each round and can invest any portion of it (e.g. \$10) with a second player. Then the investment triples,

and the second player decides how much to repay (e.g. returning \$13 and keeping \$17). Cooperation, in which higher amounts are invested and then paid back, benefits both sides but carries the risk of exploitation. In both the PDG and the trust game, the amount of striatal activity relates to greater cooperation or reciprocity in subsequent rounds (Rilling et al. 2002; King-Casas et al. 2005). Unreciprocated cooperation in the PDG was associated with increased anterior insula activity (Rilling et al. 2008), a brain region known to be associated with emotion including responses to aversive stimuli (Dayan and Seymour 2008).

Even a task without monetary rewards can show reward-related activity for cooperation. An example is a study using a computer game that involved arranging a visual pattern, which people either undertook in cooperation with another, in competition with another or alone. Cooperation led to greater activity in OFC than competition (Decety et al. 2004).

Other studies have looked at the neural processing related to reputations acquired in such games. Encountering those who had gained a reputation for cooperation in a PDG also elicits activity in reward-related ventral striatum and OFC (Singer et al. 2004). Further, when men were scanned whilst watching electric shocks administered to those they had previously played in a PDG, this led to reward-related activity when defectors were shocked but led to empathy-related responses in pain-related areas when cooperators were shocked (Singer et al. 2006).

A further study examined brain activity during a trust game, in which participants learned the reputations of others who were more or less cooperative (Phan et al. 2010). As before, participants’ ventral striatum and OFC were engaged by positive reciprocity from others. Interestingly, here this signal in ventral striatum was seen when interacting with partners who had gained a reputation for reciprocity, but absent for partners without a reputation for reciprocity. The authors suggest this reflects a mechanism involving reward-related brain regions, which initiates and sustains cooperative relationships.

In summary, contrary to the expectation from influential models that suggest humans are only

self-interested, the evidence presented here is consistent with the idea that cooperation also itself engages reward mechanisms in the brain. We next ask how the balance between the drive to cooperate and more self-interested motivations is managed over time.

Managing the Balance Between Cooperation and More Self-Orientated Behaviours

The success of social animals, particularly humans, depends on how well individuals manage a critical day-to-day trade-off between cooperative and more self-motivated behaviours. Biological mechanisms controlling this trade-off must tune behaviour to the social environment.

Because of the dominant conception from RCT that humans are only self-interested, much research has focused on identifying factors that increase a propensity to cooperate. As described above, cooperative behaviours are thought to co-opt neural reward mechanisms (Phan et al. 2010). Evidence also suggests such behaviours are causally promoted by the peptide hormone oxytocin, which has various important roles in humans and has been administered to human participants in a variety of studies (MacDonald and MacDonald 2010). For example, oxytocin increased cooperation within groups in a PDG (De Dreu et al. 2010) and also increased measures of trust in a trust game (Kosfeld et al. 2005).

However, without opposing factors, this form of control mechanism would be lopsided. Testosterone has been shown as such an opponent endocrine influence, which promotes more self-orientated behaviour and reduces cooperation (Wright et al. 2012). This gonadal hormone is secreted in men and women and modulates a range of behavioural trade-offs in humans and other animals, for example, the trade-off between parenting and courtship (Wingfield et al. 1990; Alvergne et al. 2009). Administering testosterone selectively and causally disrupted cooperation by increasing egocentricity in decision-making, operationalised as an enhanced weighting of one's own relative to another's evidence (Wright

et al. 2012). We can also see related function in non-human primates, for example, where before competitive interactions between the self and others, anticipatory testosterone rises are seen in chimpanzees but not in more cooperative and egalitarian bonobos (Wobber et al. 2010).

These hormonal influences also illustrate an advantage of a biologically based approach. One way to improve the assumptions of game theory is to invoke the concept of "other-regarding preferences" (Fehr and Camerer 2007). For example, in a game between me and you, my utility function (i.e. what I value) would include not only what I personally receive but also what you receive (weighted in some fashion). This approach can be useful, for example, providing quantified metrics on a trial-by-trial basis for use in neuroimaging analyses involving a model-based approach as described for the ultimatum game below (Wright et al. 2011). However, without the addition of enormous complexity, such models cannot explain critical features of social behaviour that can be comfortably accommodated by a biological perspective. An example is our knowledge of the endocrine system (e.g. oxytocin and testosterone above), which helps explain how the trade-off between social and self-interested motivations is dynamically modulated in response to environmental contingencies, which is critical for success of social animals such as humans.

Finally, we can look in more detail at an interesting brain imaging study that examined how humans maintain and repair breakdowns in cooperation in the trust game (King-Casas et al. 2008). When collaboration falters and investments are low, individuals often build cooperation by making unilateral conciliatory gestures in the form of high repayments, even though these may be taken and not reciprocated. These gestures are precisely tracked in individuals' anterior insula cortex, a brain region that processes important emotional responses. Successful resolution of breakdowns in negotiation can be one of the most influential means for transforming a conflict (Galluccio 2011:225). Humans use such cooperative gestures as one tool to manage the critical balance between cooperative and more self-orientated motivations.

International Negotiations

We now illustrate accommodative signals in international negotiation. During the China–US crisis over Taiwan in 1958, the United States used a combination of positive inducements as well as military stick (Spangler 1991). During the crisis, Secretary of State John Foster Dulles, a very tough operator, made accommodative gestures: firstly the accommodative signal of wish to resume talks and most notably 3 weeks later when he disavowed any commitment for a Nationalist return to the mainland and hinted at future troop reduction on the islands. These accommodative gestures, each subsequently reciprocated by mainland China, were central to resolution of the crisis.

A contemporary example is the election of Iranian President Hassan Rouhani in 2013. This followed almost a decade of near- ceaseless hostility with Western powers and reflected the desire for accommodation amongst the Iranian people. Rouhani’s pragmatism distinguished him from his more ideological competitors during the presidential campaign. Regarding negotiations with the West over Iran’s nuclear program, discussed further below, he asserted in one presidential debate: “It is good to have centrifuges running, provided people’s lives and livelihoods are also running” (Wright and Sadjadpour 2014).

Policy Recommendation

Expect accommodative and conciliatory gestures as natural and common. Do not mistake others’ positive gestures for weakness.

Fairness

A second social motivation for which there is good concordant behavioural and neural evidence is fairness. This social motivation matters because humans are prepared to pay high costs to reject unfairness. Fairness relates to how intentional agents should divide resources amongst potentially entitled recipients (Kahneman et al.

1986) and has interested economists (Akerlof 1979), sociologists (Homans 1961), as well as neuroscientists (Sanfey et al. 2003).

A classic illustration of fairness is the ultimatum game (UG). In the UG one player (the proposer) is given an endowment (e.g. £10) and proposes a division (e.g. keep £6/offer £4) to a second player (the responder), who can accept (both get the proposed split) or reject (both get nothing) the offer (Güth et al. 1982). Game theory predicts that if individuals maximise only their own payoffs, then responders should accept any offer (1 penny is better than nothing) and, knowing this, proposers should offer as little as possible.

Instead, humans are prepared to pay a high cost to reject unfairness and reject offers below 25 % about half the time (Camerer 2003). This has been shown across diverse cultures (Henrich et al. 2006) and with large stakes (Slonim and Roth 1998; List and Cherry 2000; Andersen et al. 2011). Further, even in a version of the UG with the responder’s ability to reject the offer removed (called a dictator game), proposers still do not offer zero, suggesting that “fair-minded” behaviour is not only due to fear of rejections (Camerer 2003).

Neural Bases of the Fairness Motivation

Neurally, considerable work links the insula cortex to the fairness motivation (Fig. 5.2). Within insula cortex, distinct fairness-related processes appear to be expressed in segregated regions (Wright et al. 2011) of this extensive (over 5 cm long) and cytoarchitectonically diverse brain region (Flynn 1999; Varnavas and Grand 1999). We can consider posterior insula, the part more towards the back of the head, and anterior insula that is more towards the front.

In the UG, in each trial a precise measure of inequality can be calculated (e.g. an 8:1 split would have an inequality of 7)—and neural activity in posterior insula negatively correlated with this measure of inequality (Wright et al. 2011). The same negative correlation with inequality in

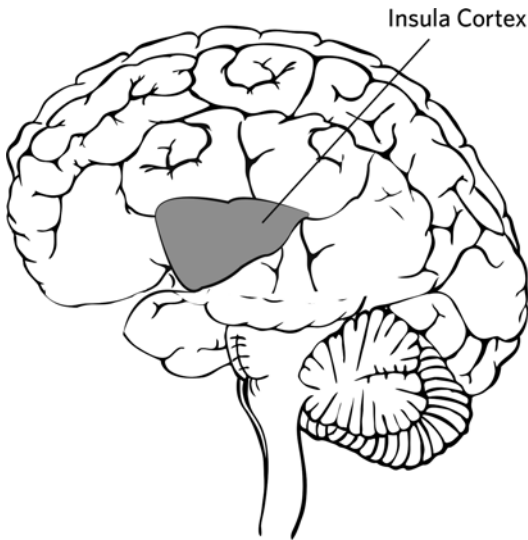


Fig. 5.2 Insula cortex is a large and diverse region that serves a number of functions, including important emotional responses

posterior insula was also seen in a very different task, in which participants chose between distributions of meals for African children that varied in inequality (measured in this case by the Gini coefficient) and amount (see Fig. 4 in Hsu et al. 2008). These concordant neural findings are striking, as Hsu et al. used decisions about third parties rather than first-party decisions (e.g. the UG in Wright et al. 2011), a difference known to markedly affect choice in behavioural experiments (Camerer 2003).

However, whilst posterior insula activity negatively correlated with inequality, anterior insula activity positively correlated with inequality (Sanfey et al. 2003), a result replicated in a task-matched study (Halko et al. 2009). This increased anterior insula activity for more unfair offers was related by the authors to moral “disgust” at the unfair offers, in light of the region’s role in processing disgust more broadly (Sanfey et al. 2003).

Since these human fMRI studies, a causal study in non-human primates using stimulation in insula has shown results highly consistent with this segregation (Caruana et al. 2011). As described above, in the human studies, posterior insula negatively correlated with inequality or put another way showed increased activity with

more prosocial behaviours (Hsu et al. 2008; Wright et al. 2011), whilst anterior insula positively correlated with inequality (Sanfey et al. 2003). Applying electric current to stimulate more posterior regions of insula led to affiliative behaviours, whilst stimulation more anteriorly led to more disgust-related behaviours (Caruana et al. 2011).

In addition to insula cortex, reward-related brain regions have also been associated with the fairness motivation in decision-making. Fair treatment in the UG has been linked with reward-related activity, where comparing fair offers with unfair offers of equal monetary value showed increased activity in regions including striatum and ventromedial prefrontal cortex (vmPFC, a reward-related region next to OFC) (Tabibnia et al. 2008). Patients with lesions to vmPFC are more likely than control subjects to reject low offers in the UG (Koenigs and Tranel 2007). In tasks outside the UG, striatum and vmPFC showed greater activity for inequality-reducing wealth transfers in a task where subjects rated wealth transfers to themselves or another individual, one of whom at the beginning of the experiment was randomly rendered “rich” and the other “poor” (Tricomi et al. 2010).

Finally, we note behavioural evidence in non-human primates of rejection of unequal treatment. In a well-known example, when two capuchin monkeys were instructed to carry out the same task and one received tasty grape whilst the other received humdrum cucumber, there was rejection of the latter food (Brosnan and De Waal 2003).

Fairness in International Negotiations

The motivation to reject unfairness and the humiliation from unfair treatment can form a central part of national narratives and are reflected in national decision-making. In a powerful Chinese narrative, “unequal treaties” in the nineteenth century with external powers, mostly Western, unfairly exploited China’s weakness, leading to a “century of humiliation” (Wang 2012). This instils a sense of entitlement to

recover and receive restitution for past losses. This played into the Chinese border clash with the Soviet Union in 1969, where scores died on both sides and nuclear threats were levelled (Gerson 2010). The Chinese were motivated in part by the desire to revise one of the old unequal treaties with Russia—the 1860 Treaty of Peking, for which the Soviets had refused the Chinese request 4 years before to recognise as an unequal treaty. And the specific objection was how to split the uninhabited, useless islands in the river Ussuri between the two countries: the Soviets wanted them all, the Chinese an equal split. It was the Chinese who initiated the military confrontation despite overwhelming Soviet nuclear and local conventional superiority.

Robert Shiller and George Akerlof, both recent Nobel laureates in economics, show how fairness shapes our national economies, for example, being central to wage negotiations (Akerlof and Shiller 2009). International economics is also affected. In 2003 World Trade negotiations, countries like Brazil walked away from a deal in which they felt developed nations did not give up enough, even at the cost of giving up gains for themselves (Kapstein 2008).

Iran has been prepared to reject perceived unfairness even at substantial cost. In 1951, Iranian Prime Minister Mohammed Mossadegh, rather than accede to an inequitable 10–90 oil deal with the British-run Anglo-Iranian Oil Company, subjected his country to a crippling embargo and a British-American-aided coup that brought about his demise. Contemporary Iran has not been deterred from continuing to develop its nuclear programme, despite costs over \$100 billion (Wright and Sadjadpour 2014). As Iranian Foreign Minister Javad Zarif asked in a YouTube message during the nuclear negotiations: “Imagine being told that you cannot do what everyone else is doing. Would you back down? Would you relent? Or would you stand your ground?” (Zarif 2013). From an Iranian perspective, the situation is one where six global powers who together possess thousands of nuclear weapons seek to dictate terms to Iran, and India and Pakistan did not sign the nuclear proliferation treaty (NPT) and secretly acquired nuclear

weapons but are accepted by the international community whilst Iran (an NPT signatory) is chastised. This impulse to reject perceived unfairness arguably motivated Iran’s nuclear ambitions far more than an actual desire or need for nuclear energy (Wright and Sadjadpour 2014).

Fairness also shapes possible deals and political necessities. First, in contemporary Iranian nuclear negotiations consider the “right” to enrich. It is hard to explain convincingly to an Iranian why Iran isn’t allowed to do something its neighbours—India, Pakistan and Israel—can do. Iran has been, and will continue to be, prepared to pay heavily to reject this inequality (Wright and Sadjadpour 2014). Any viable agreement will likely enable Iranians to say they have that right, even if the word isn’t in the text. Second, the social motivation can shape the specific form of events during a crisis. For example, in 2001 a US EP-3 reconnaissance plane and a Chinese fighter collided, which led to the loss of the Chinese pilot and forced the US plane to land on Hainan in China. The key Chinese demand was for an apology (Swaine et al. 2006).

Policy Recommendations: Fairness

1. Use knowledge of this motivation to understand intentions and so build a better account of the other. The injunction to look from the other’s perspective is a very broad recommendation—and understanding this social motivation gives a targeted question: “Was this seen as fair or unfair?” This helps explain key facts, e.g.: Why has contemporary Iran borne costs estimated at \$100 billion to pursue its nuclear programme? Why does China care so much about territory related to the unequal treaties and associated events? Training for negotiators and mediators can include cognitive, emotional and motivational insights to understand intentions and behaviours (Aquilar and Galluccio 2008).
2. Forecasting the other’s decision calculus: These forecasts can be incorrect without incorporating the value of unfairness. To correctly understand another’s decision calculus,

we must consider social motivations. Ask the targeted question: “Is this seen as fair or unfair?” Consider the Sino-Soviet border conflict described above, where there was a failure of deterrence despite massive Soviet conventional and nuclear superiority—the Soviets incorrectly forecast the Chinese decision calculus. Consider a China-US escalation scenario: when the Chinese deal with the Japanese over territorial issues, it may take more to deter the Chinese than might otherwise be understood.

3. Know how fairness shapes possible deals: Anticipate these political realities, such as in the descriptions above of contemporary Iranian nuclear negotiations and Sino-US crisis management. This helps you understand what the other side values highly that you may not value so highly, enabling you to make a favourable trade.

The Neural Phenomenon of “Prediction Errors” Exerts Impacts Throughout Diplomatic and Military Signalling

Finally, to manage negotiations, it is necessary to forecast how the other will decide to respond to our actions. Consider the situation where the other has made an action to which we must respond. How do we implement a calibrated response? To exert our intended degree of impact on their decision-making, we must understand how the psychological impact of actions is modulated by a key quantity in the brain’s decision-making circuits. This quantity is the difference between what happens and what was expected. It is called “prediction error”. The prediction error associated with an event modulates the event’s impact on decision-making, and the bigger the prediction error, the bigger the impact. We must understand prediction errors to forecast the impact of our actions on others—and they provide a simple, powerful tool.

Prediction errors are best understood neuroscientifically in the case where animals and humans get rewards or punishments (Schultz et al. 1997; O’Doherty et al. 2004), but the broader idea is

	Event not expected	Event expected
Event occurs	a) Associated with prediction error	b) No prediction error
Event not occur	(trivial case)	c) Associated with prediction error

Fig. 5.3 Illustrating prediction errors

involved in many neuroscientific models (Friston 2010). From **simple tasks** (Niv and Schoenbaum 2008) to more complex **social** interactions (Behrens et al. 2009), it is central to how humans understand, learn and decide about the world. (Note this section draws on Wright, 2014).

Signalling Between Nations

Prediction errors exert far-reaching impacts, and these can be captured by a simple framework. Considering a simple definition of prediction error as the difference between what happened and what was expected (i.e. $prediction\ error = actual\ event - expected\ event$). This gives a simple framework: the event can either occur or not occur and either be expected or not expected (Fig. 5.3).

A dramatic illustration of the three non-trivial types of event in Fig. 5.3 is given by the psychological impact of strategic bombing during wartime (Quester 1990; Lambert 1995). First consider an event that occurs and was not expected, so has a large associated prediction error (Fig. 5.3a). German air raids on London in the First World War using zeppelins were small scale, but being so unexpected, they had a large impact and caused panic.

Between the wars, highly influential airpower theorists like Douhet extrapolated from this to suggest that more powerful and recurrent bombing would, largely through psychological impact, paralyse adversaries and rapidly make them collapse. But what actually happened illustrates an

event that occurs but is well expected (Fig. 5.3b). In the Second World War, recurrent bombing exerted much greater destructive power, for example, the “Blitz” on London, but being expected it had much more limited psychological impact than forecast.

Third, an event is expected but doesn’t occur, so the absence of a predicted event leads to large prediction error (Fig. 5.3c). In the Vietnam War, during regular US bombing of North Vietnam, the United States used prolonged bombing pauses as a conciliatory signal.

The cases above involve punishing events, but prediction errors equally apply to conciliatory acts. Consider the actions of Egyptian leader Anwar Sadat in 1977. Egypt had lost two wars to Israel in 1967 and 1973, after which he made conciliatory efforts that did not markedly change the attitudes of Israeli decision-makers or public (Mitchell 2000). However, in 1977 he made the highly unexpected novel offer to go and speak in the Israeli Knesset—and this had a big psychological impact on both Israeli decision-makers and the public and opened the path to reconciliation (Mitchell 2000).

We can also consider the nuclear negotiations with contemporary Iran in late 2013 (Wright and Sadjadpour 2014). A number of unexpected gestures helped create the opportunity for the negotiations. In 2009 there was US President Obama’s unexpected video to the Iranian people and “leadership of the Islamic Republic of Iran” and two unprecedented private letters to Iranian Supreme Leader Khamenei. These overtures helped persuade the Iranian public of America’s interest in change. In September 2013, there was the unexpected “Twitter diplomacy” of newly elected Iranian President Rouhani and Javad Zarif, which shifted the tone of America’s foreign policy debate about Iran. Then in September 2013, there was the unprecedented Obama-Rouhani phone call during the UN General Assembly, which built confidence in both countries.

Finally, we note that a prediction error framework subsumes and explains core concepts in negotiation. For example, the psychological impact of surprise is an instance of prediction error, where an event has occurred

but is not well predicted (Fig. 5.3a). It also encompasses other concepts, including habituation, expectation management, learning and adaptability, and signposting.

Policy Recommendations

We can consider policy recommendations first when making actions and second when receiving actions.

Making Actions

The core idea is to use prediction errors as a tool in signalling.

1. *When preparing potential options for a decision-maker, for each option ask: “How unexpected will it be for the other?”* For each option describe its associated prediction error from the other’s perspective and how that modulates its signalling impact.
2. *Manipulate predictability.* The other side of the coin of prediction error or surprise is predictability. This manipulates the signalling impact of actions, e.g. signpost or telegraph actions.

Receiving Actions

The core idea is that prediction errors are unavoidable, so we must manage their effects on oneself.

1. *Manage effects of prediction errors:* Prediction error may lead to a large psychological impact on decision-makers and they should be aware of this so they react appropriately.
2. *Learning:* Prediction errors are the best material to improve our models of the world and our models of the other.

Discussion and Conclusion

Biological and neuroscientifically based approaches to choice have a long theoretical and empirical tradition (Thorndike 1911; Mackintosh 1983)—and have more recently been combined with economics and psychology to provide an extra source of evidence about decision-making. Above I gave three insights from the neuroscien-

tifically grounded account of choice that help us forecast how an adversary will decide to respond to our actions. Next, I describe four general rules (Wright, 2013) for using neuroscience, and the behavioural decision sciences more generally, to address practical policy issues.

First, are we sure enough of the neuroscience? In a rapidly advancing field like neuroscience, there are a plethora of ideas and findings. For this reason I focused on robust findings.

Second, does it matter in the real world? Such findings may be very convincing in individuals making particular decisions, for example, in a lab—but in the real world, with all its complexities and existing structures and unintended or unpredictable consequences, we may not see such an effect. Here I adopt a similar approach to the seminal work of Robert Jervis who applied insights from psychology to international relations (Jervis 1976). Specifically, here I use perspectives from a neuroscientifically grounded account of decision-making and show how they explain a variety of historical cases across different contexts. With respect to how these aspects of individual decision-making affect international negotiation, they may directly affect decision-makers themselves and/or shape the reactions of the public or key interest groups and so influence the political landscape in which the decision-makers must operate.

Third, even if it is true in the real world, is it worth adding to the policy process? Given all the many important considerations when developing or using policy, adding yet another consideration can carry a big opportunity cost. Here, for instance, instead of adding to the analytic burden faced by decision-makers and their staff, the prediction error framework described above replaces and simplifies across a wide range of important phenomena.

Fourth, what does the neuroscience add that behavioural approaches, such as psychology or economics, do not already give us? There is the important concept of “consilience” (Wilson 1999): psychology is only one source of evidence to explain behaviour, and we can be more confident of a particular explanation if it is supported by both psychological and neuroscientific evidence. Neuroscience can help choose between otherwise similarly plausible behavioural explanations, by

looking in the brain for parts of the mechanism proposed to underlie behaviour (O’Doherty et al. 2007). Further, a robust biological basis for a decision-making behaviour enhances our prior belief about the generalisability of findings across cultures, which is crucial in international negotiations—if we know prediction errors play an important role in decision-making across a wide variety of different species, including in humans, then it is much more likely that they play an important role in, for example, both the United States and China. A biological perspective also helps improve our prior beliefs about generalisability within countries or cultures, for example, as key policymakers have usually undergone an involved selection process and so may differ from the general population. No single approach—including neuroscience, psychology or economics—explains human decision-making, and neuroscience provides an important extra source of evidence.

I have presented three insights from the rapidly advancing field that combines neuroscience, psychology and economics. I have also provided historical examples and practical policy recommendations. This new approach helps provide a robust explanation of human motivation and decision-making in international negotiation.

References

- Akerlof, G. A. (1979). The case against conservative macroeconomics: An inaugural lecture. *Economica*, *46*, 219–237.
- Akerlof, G. A., & Shiller, R. J. (2009). *Animal spirits*. Princeton, NJ: Princeton University Press.
- Alvergne, A., Faurie, C., & Raymond, M. (2009). Variation in testosterone levels and male reproductive effort: Insight from a polygynous human population. *Hormones and Behavior*, *56*, 491–497.
- Andersen, S., Ertac, S., Gneezy, U., Hoffman, M., & List, J. A. (2011). Stakes matter in ultimatum games. *American Economic Review*, *101*, 3427.
- Aquilar, F., & Galluccio, M. (2008). *Psychological processes in international negotiations*. New York: Springer.
- Behrens, T. E. J., Hunt, L. T., & Rushworth, M. F. S. (2009). The computation of social behavior. *Science*, *324*, 1160–1164.
- Brosnan, S. F., & de Waal, F. B. M. (2002). A proximate perspective on reciprocal altruism. *Human Nature*, *13*, 129–152.

- Brosnan, S. F., & De Waal, F. B. (2003). Monkeys reject unequal pay. *Nature*, *425*, 297–299.
- Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.
- Camerer, C. F., & Loewenstein, G. (2004). Behavioral economics: Past, present, future. In C. F. Camerer, G. Loewenstein, & M. Rabin (Eds.), *Advances in behavioral economics* (pp. 3–51). Princeton, NJ: Princeton University Press.
- Camerer, C., Loewenstein, G., & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, *43*, 9–64.
- Caruana, F., Jezzini, A., Sbriscia-Fioretti, B., Rizzolatti, G., & Gallese, V. (2011). Emotional and social behaviors elicited by electrical stimulation of the insula in the macaque monkey. *Current Biology*, *21*, 195–199.
- Dayan, P., & Seymour, B. J. (2008). Values and actions in aversion. In *Neuroeconomics: Decision making and the brain*. London: Academic Press.
- De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., Van Kleef, G. A., Baas, M., Ten Velden, F. S., Van Dijk, E., & Feith, S. W. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science*, *328*, 1408–1411.
- Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: An fMRI investigation. *NeuroImage*, *23*, 744–751.
- Dugatkin, L. A. (1997). *Cooperation among animals: An evolutionary perspective*. Oxford: Oxford University Press.
- Fehr E, Camerer CF (2007) Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences* *11*, 419–427.
- Flynn, F. G. (1999). Anatomy of the insula functional and clinical correlates. *Aphasiology*, *13*, 55–78.
- Frackowiak, R. S. J., Ashburner, J. T., Penny, W. D., Zeki, S., Friston, K. J., Frith, C. D., Dolan, R. J., & Price, C. J. (Eds.). (2004). *Human brain function* (2nd ed.). San Diego, CA: Academic Press.
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138.
- Galluccio, M. (2011). Transformative leadership for peace negotiation. In F. Aquilar & M. Galluccio (Eds.), *Psychological and political strategies for peace negotiation: A cognitive approach*. New York: Springer.
- Gerson, M. S. (2010). The Sino-Soviet border conflict. CNA. <http://www.cna.org/research/2010/sino-soviet-border-conflict>.
- Glimcher, P. W. (2004). *Decisions, uncertainty, and the brain: The science of neuroeconomics* (1st ed.). Cambridge, MA: The MIT Press.
- Glimcher, P. W., & Fehr, E. (2013). *Neuroeconomics decision making and the brain*. Amsterdam, MA: Academic Press.
- Glimcher, P. W., & Rustichini, A. (2004). Neuroeconomics: The consilience of brain and decision. *Science*, *306*, 447–452.
- Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*, 367–388.
- Halko M-L, Hlushchuk Y, Hari R, Schürmann M (2009) Competing with peers: mentalizing-related brain activity reflects what is at stake. *Neuroimage* *46*, 542–548.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*, 1767–1770.
- Ho, T. H., Lim, N., & Camerer, C. F. (2006). Modeling the psychology of consumer and firm behavior with behavioral economics. *Journal of Marketing Research*, *43*, 307–331.
- Homans, G. C. (1961). *Social behavior: Its elementary forms*. New York: Harcourt, Brace & World.
- Hsu, M., Anen, C., & Quartz, S. R. (2008). The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science*, *320*, 1092–1095.
- Jervis, R. (1976). *Perception and misperception in international politics*. Princeton, NJ: Princeton University Press.
- Kagel, J. H., & Roth, A. E. (1995). *The handbook of experimental economics*. Princeton, NJ: Princeton University Press.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, *47*(2), 263–291.
- Kahneman D, Knetsch JL, Thaler R (1986) Fairness as a Constraint on Profit Seeking: Entitlements in the Market. *The American Economic Review* *76*, 728–741.
- Kapstein, E. B. (2008). Fairness considerations in world politics: Lessons from international trade negotiations. *Political Science Quarterly*, *123*, 229–245.
- King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science*, *321*, 806–810.
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, *308*, 78–83.
- Koenigs, M., & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the ultimatum game. *Journal of Neuroscience*, *27*, 951–956.
- Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, *435*, 673–676.
- Lambert, A. P. N. (1995). *The psychology of air power*. London: Royal United Services Institute for Defence Studies.
- List, J. A., & Cherry, T. L. (2000). Learning to accept in ultimatum games: Evidence from an experimental design that generates low offers. *Experimental Economics*, *3*, 11–29.
- MacDonald, K., & MacDonald, T. M. (2010). The peptide that binds: A systematic review of oxytocin and its

- prosocial effects in humans. *Harvard Review of Psychiatry*, 18, 1–21.
- Mackintosh, N. J. (1983). *Conditioning and associative learning*. New York: Clarendon.
- Mitchell, C. R. (2000). *Gestures of conciliation: Factors contributing to successful olive branches*. New York: St. Martin's Press.
- Niv, Y., & Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends in Cognitive Sciences*, 12, 265–272.
- O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Current Opinion in Neurobiology*, 14, 769–776.
- O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, 304, 452–454.
- O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences*, 1104, 35–53.
- Phan, K. L., Sripada, C. S., Angstadt, M., & McCabe, K. (2010). Reputation for reciprocity engages the brain reward center. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 13099–13104.
- Quester, G. H. (1990). The psychological effects of bombing on civilian populations: Wars of the past. In B. Glad (Ed.), *Psychological dimensions of war* (pp. 201–214). Newbury Park, CA: Sage.
- Rilling, J. K., Goldsmith, D. R., Glenn, A. L., Jairam, M. R., Elfenbein, H. A., Dagenais, J. E., Murdock, C. D., & Pagnoni, G. (2008). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia*, 46, 1256–1266.
- Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron*, 35, 395–405.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300, 1755–1758.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275, 1593–1599.
- Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron*, 41, 653–662.
- Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439, 466–469.
- Slonim, R., & Roth, A. E. (1998). Learning in high stakes ultimatum games: An experiment in the Slovak republic. *Econometrica*, 66, 569–596.
- Spangler, S. E. (1991). *Force and accommodation in world politics*. DIANE Publishing. Maxwell AFB, Ala.: Air University Press.
- Swaine, M. D., Zhang, T., & Cohen, D. F. S. (2006). *Managing Sino-American crises: Case studies and analysis*. Washington, DC: Carnegie Endowment for International Peace.
- Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science*, 19, 339–347.
- Thorndike, E. L. (1911). *Animal intelligence: Experimental studies*. New York: Macmillan.
- Tricomi, E., Rangel, A., Camerer, C. F., & O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature*, 463, 1089–1091.
- Varnavas, G. G., & Grand, W. (1999). The insular cortex: Morphological and vascular anatomic characteristics. *Neurosurgery*, 44, 127–136. discussion 136–138.
- Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.
- Wang, Z. (2012). *Never forget national humiliation: Historical memory in Chinese politics and foreign relations*. New York: Columbia University Press.
- Wilson, E. O. (1999). *Consilience: The unity of knowledge*. London: Abacus.
- Wingfield, J. C., Hegner, R. E., Dufty, A. M., & Ball, G. F. (1990). The “challenge hypothesis”: Theoretical implications for patterns of testosterone secretion, mating systems, and breeding strategies. *The American Naturalist*, 136, 829–846.
- Wobber, V., Hare, B., Maboto, J., Lipson, S., Wrangham, R., & Ellison, P. T. (2010). Differential changes in steroid hormones before competition in bonobos and chimpanzees. *Proceedings of the National Academy of Sciences*, 107, 12457.
- Wright, N. D., Bahrami, B., Johnson, E., Di Malta, G., Rees, G., Frith, C. D., & Dolan, R. J. (2012). Testosterone disrupts human collaboration by increasing egocentric choices. *Proceedings of the Royal Society B: Biological Sciences*, 279, 2275–2280.
- Wright (2013) in Proceedings of 7th Annual Conference, *Strategic Multilayer Assessment, Joint Staff DoD*.
- Wright, N. D., & Sadjadpour, K. (2014). The neuroscience guide to negotiations with Iran. *The Atlantic*. Available at: <http://www.theatlantic.com/international/archive/2014/01/the-neuroscience-guide-to-negotiations-with-iran/282963/>. Accessed March 12, 2014.
- Wright Nicholas D., Neural prediction error is central to diplomatic and military signalling (2014) in DiEuliis D, Casebeer W, Giordano J, Wright ND, Cabayan H (Eds) White paper on Leveraging Neuroscientific and Neurotechnological (NeuroS&T) Developments with Focus on Influence and Deterrence in a Networked World, US DoD Joint Staff.
- Wright, N. D., Symmonds, M., Fleming, S. M., & Dolan, R. J. (2011). Neural segregation of objective and contextual aspects of fairness. *Journal of Neuroscience*, 31, 5244–5252.
- Zarif, J. (2013). *Iran's message: There is a way forward*. Available at: http://www.youtube.com/watch?v=Ao2WH6GDWz4&feature=youtube_gdata_player. Accessed March 21, 2014.