

# Journal of Theoretical Politics

<http://jtp.sagepub.com/>

---

## **Rationality, Inaccurate Mental Models, and Self-confirming Equilibrium : A New Understanding of the American Revolution**

Rui J. P. de Figueiredo, Jr., Jack Rakove and Barry R. Weingast

*Journal of Theoretical Politics* 2006 18: 384

DOI: 10.1177/0951629806067451

The online version of this article can be found at:

<http://jtp.sagepub.com/content/18/4/384>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *Journal of Theoretical Politics* can be found at:**

**Email Alerts:** <http://jtp.sagepub.com/cgi/alerts>

**Subscriptions:** <http://jtp.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://jtp.sagepub.com/content/18/4/384.refs.html>

>> [Version of Record](#) - Sep 6, 2006

[What is This?](#)

## RATIONALITY, INACCURATE MENTAL MODELS, AND SELF-CONFIRMING EQUILIBRIUM

### A NEW UNDERSTANDING OF THE AMERICAN REVOLUTION

Rui J. P. de Figueiredo Jr., Jack Rakove and  
Barry R. Weingast

#### ABSTRACT

Rational choice theorists have long been ambivalent about the rationality postulate. Although many agree that humans have imperfect foresight and inaccurate understandings of the world within which they act, no satisfactory approach relaxing the rationality postulates exists. In this article, we apply a new method that allows us to model directly rational action when agents have incomplete or inaccurate mental models of their world. In particular, we utilize the game theoretic notion of *self-confirming equilibrium*. The self-confirming equilibrium framework allows us to model an environment where a player's predictions about one another's behavior are inaccurate. It also allows us to model a common aspect of disputes; *viz.*, that both sides believe they are acting reasonably while the other side is acting unreasonably. Finally, the approach allows a type of surprise in which a player's mis-conjectures about another's behavior results in behavior that was not anticipated. We apply this framework to the problem of the American Revolution, helping to resolve several unexplained puzzles in the historical literature. Historians of the Revolution emphasize the role of *ideas* in underpinning the revolutionary crisis. From the rational choice standpoint, a critical omission in the literature is that historians do not connect the realm of ideas with the realm of action. We demonstrate how the framework we propose can more directly shed light on three aspects absent in the existing literature. First, to explain why either side fought about abstract ideals or how those ideals were connected to the realm of action, and thus why the two sides failed to come to some accommodation despite their differences. Second, to explain why the colonists became so incensed over seemingly trivial taxes imposed by the British after the Seven Years' War ended in 1763. And, finally, to explain that if the clash of fundamental ideals led the two sides to armed conflict,

---

The authors gratefully acknowledge helpful conversations with Cheryl Boudreau, Colin Camerer, Daniel Carpenter, John Ferejohn, Robert Gibbons, Joseph Greenberg, Stephen Krasner, David Laitin, Dimitri Landa, Mathew McCubbins, Paul Milgrom, James Morrow, Roger Myerson, Douglass North, Robert Powell, Eric Rasmussen, Andy Rutten, Thomas Sargent, and Robert Wilson.

how the British and the Americans could be unaware of their profound differences during the previous 100 plus years of cooperation.

**KEY WORDS** • American revolution • rationality • self-confirming equilibrium

## 1. Introduction

Rational choice theorists have long been ambivalent about the rationality postulate. Although many agree that humans have imperfect foresight and inaccurate understandings of the world within which they act, no satisfactory approach relaxing the rationality postulate exists. Theorists have taken several related tacks in their attempts to relax the boundaries of rationality. One group of scholars has attempted to incorporate specific lessons from cognitive science, psychology, or experiments that reveal limits on rational thinking. Recent examples include Camerer (1995), Denzau and North (1994), Elster (1998, 1999), Frey (1997), Frank (1985, 1989); Kahneman and Tversky (1979), Lupia and McCubbins (1998), and Rabin (1998).<sup>1</sup> A second approach reflects scholars who argue that under certain circumstances, particular mechanisms of interaction imply that cognitive limits are not binding (Satz and Ferejohn, 1994; Clark, 1997a, 1997b; see also North, 2004, drawing on Hayek, 1960). Third, a group of scholars using game theory and related approaches have explored alternative definitions of rationality that directly incorporate individuals' limited knowledge and abilities (see Fudenberg and Levine, 1993, 1998; Kalai and Lehrer, 1993; Greenberg, 1994; Fudenberg and Kreps, 1995; Sargent, 1999; Greenberg et al., 2004; Lupia and Zharinova, 2004).

In this article, we draw on this third literature. Rather than incorporating a specific type of limit on rationality, we develop a new method that allows us to model directly rational action when agents have incomplete or inaccurate mental models of their world. At times, individuals act in such great ignorance of their world that unexpected results occur at every turn. These circumstances are not the subject of this article. We instead study circumstances where mental models have typically evolved over time and where they have, in a particular sense, stood the test of time. The perspective requires an explanation of how, in the face of their ignorance, rational individuals

---

1. This list hardly exhausts the categories of work on this topic. For example, there is the long-standing work of Simon (1947) and March and Simon (1958) on bounded rationality; game theorists have studied 'finite state automata; and there exist several collections of essays where various proponents and critics of rational choice theory discuss these issues' (see e.g., Cook and Levi, 1990; and Mansbridge, 1990).

nonetheless continue to rely on their inaccurate mental models (Kahneman and Tversky, 1979). Our answer is not of the form often given by those drawing on cognitive science, such as the hypothesis that individuals have a tendency to disregard evidence that contradicts their own point of view. Instead, we follow the game theorists noted earlier by incorporating inaccuracies directly into a model.

In the real world, humans face a variety of cognitive limits. But this does not imply that they are irrational – in the sense they do not behave optimally given their interests and understanding of their situation – or that they ignore experience. Following Greenberg (1994) and Sargent (1999), we assume that people attempt to understand their world, and their theories about the world in equilibrium are consistent with their experience. Because the players are not omniscient and have limited experience, they may have incomplete or inadequate models of their environment.

In terms of the approach, we allow the players in a game to have two types of inadequate understandings of the world. First, they may not completely understand the game they are playing; and second, they may have mistaken beliefs about the strategies used by the other players. A critical feature of the approach is that, although the players can have inaccurate models of their environment, these inaccuracies cannot be arbitrary. The approach constrains inaccuracies to occur within an equilibrium framework that imposes discipline on the types of mistakes and misunderstandings. Specially, the framework requires that nothing the players observe along the equilibrium path of play disconfirms their beliefs about the game they are playing or about the strategies being played by others. Although misunderstandings and inaccuracies are possible, they must be mutually consistent in this equilibrium sense.

This approach contrasts with the typical non-rational accounts which posit that the actors misunderstand their world in particular ways. In these approaches, the errors can be arbitrary. In our approach, the errors must be consistent with experience, in particular, with observations of the other players' choices.

To study these phenomena, we employ a variant of the concept of self-confirming equilibrium, a generalization of the traditional Nash equilibrium framework in game theory (Fudenberg and Levine, 1993, 1998; Greenberg, 1994; Sargent, 1999).<sup>2</sup> In the context of extensive form games, self-confirming equilibrium relaxes the common knowledge assumption that all players have the same understanding of their strategic environment. Players instead may have very different understandings of their strategic situations. The equilibrium component imposes discipline by requiring that nothing an agent

---

2. In the study of politics, see also Lupia and Zharinova (2004).

observes disconfirms her views about her environment. The operational difference between self-confirming equilibrium and Nash equilibrium is thus twofold: self-confirming equilibrium allows the players to have certain types of incorrect conjectures about the structure of the game and about one another's strategies.

The self-confirming equilibrium framework allows us to model an environment where a player's predictions about one another's behavior are inaccurate. Indeed, a player may be ignorant of her uncertainty, a phenomenon ruled out in the traditional Nash equilibrium framework. The framework also allows us to model a common aspect of disputes; viz., that both sides believe they are acting reasonably while the other side is acting unreasonably. In the Nash equilibrium framework, such interpretations are either epiphenomenal or irrelevant.

The approach allows a type of surprise in which a player's misconceptions about another's behavior result in behavior that was not anticipated. Consider a set of players in a self-confirming equilibrium, perhaps one that has been stable for multiple generations. Suppose that an exogenous shock occurs which, given their strategies, causes one or more players to alter their preferred moves. Because these moves were off-the-path prior to the exogenous shock, the players had no prior experience with this environment. Prior to the shock, their beliefs about one another's strategy may have been inaccurate, but those inaccuracies are never discovered and therefore never invalidated. Under these circumstances, one or more players may surprise the others by making an unanticipated choice when the shock occurs.

We apply our framework to the problem of the American Revolution, helping to resolve several unexplained puzzles in the historical literature. Historians emphasize the role of *ideas* in underpinning the revolutionary crisis of 1763–76.<sup>3</sup> These historians focus on the different ways in which the Americans and the British interpreted the principal issues. The British and the Americans profoundly disagreed about the nature of the constitution governing the exercise of political authority within the empire. Specific issues included the extent of British authority within the colonies, the location of sovereignty, and the structure of the empire. Also central to the dispute was a debate about the mechanisms protecting liberty in America.

From the rational choice standpoint, a critical omission in this historical literature is its difficulty in connecting the realm of ideas with the realm of action. Indeed, historians fail to explain a range of questions about the American Revolution (see Rakove et al., 1999). For our purposes, we consider three puzzles. First, historians do not adequately explain why either

---

3. Major works include: Morgan (1956), Bailyn (1967), Greene (1986), Rakove (1979, 1996), Reid (1995) and Wood (1969).

side fought about abstract ideals or how those ideals were connected to the realm of action, and thus why the two sides proved unable to accommodate their differences. Second, why did the colonists become so incensed over the relatively modest taxes the British tried to impose after the Seven Years' War ended in 1763? Finally, given that the clash of fundamental ideals ended in armed conflict, how could the British and the Americans have been unaware of their profound differences during the previous century of cooperation? A satisfactory account for the Revolution must explain why the two sides could cooperate for so long despite their differences, only to drift into war in a few short years.

Along with a companion paper (Rakove et al., 1999), we here venture new answers to these questions. This argument begins with the theory of credible commitments to interpret how American leaders argued that the traditional institutions of the empire credibly protected their rights. During the revolutionary crisis, Americans developed a theory about how a range of institutions provided credible commitment to their rights and liberties. Relevant institutions included the structure of the empire, the division of authority between colonies and metropole, and the role of colonial assemblies in protecting a range of American liberties. Indeed, without these commitments, the pre-Revolution period – one in which a stable, cooperative equilibrium prevailed – may have looked much different; it was the assurances that American perceived that led to acquiescence in empire.

Using the concept of credible commitments, we provide the critical link between the ideals expressed by the Americans during the revolutionary debates and the realm of politics, economics, and everyday life. Alarm rose across the colonies as many Americans came to believe that the British were abandoning their long-term credible commitments to American security, putting American ways of life and sources of livelihoods at risk. Further, British interventions in the colonies threatened the view that British intentions toward America were benevolent (Rakove et al., 1999).

Using the self-confirming equilibrium perspective, our account helps interpret these claims and to address the above puzzles. Turning to the ideas themselves, the Americans looked back to the English ideas of the late 17th century to understand the constitution of the empire (Greene, 1986; Reid, 1995). According to this perspective, long-established practice and custom were accorded constitutional status. Americans also interpreted the empire as having a federal structure (though they did not use this label). The British managed the system-wide policies, such as trade and security, while the Americans retained control over their domestic affairs, including taxation, religion, and commerce. This control also covered property rights in a range of assets, such as land and slaves, and thus the principal sources of livelihood for most Americans. The interpretation of empire as federal

allowed the Americans to infer that sovereignty within the empire was divided – indeed, that it had been for over 100 years.

Taken together, these two principles had a critical implication: the American perspective indicated that the empire's federal structure, enshrined in several generations of practice, had attained constitutional status. According to this interpretation, the British lacked authority to regulate or otherwise control domestic colonial affairs. Any attempt by the British to do so violated the constitution and was therefore arbitrary. The taxes proposed by the British may have been trivial in themselves, but if the British had authority to impose that tax, revolutionary theorists argued, they also had authority to alter all colonial laws and rights and thus to confiscate American wealth.

The British interpreted these issues very differently. By the latter half of the 18th century, the British had begun to develop a new view of the constitution that matured in the 19th century. This view held that Parliament was the supreme authority in Britain and that it had the power to alter the constitution. The British also believed that the structure of the empire was a matter of policy choice having no constitutional status. Allowing the colonists control over their domestic affairs was a privilege, not a right. The fact that the British had long abstained from exercising their powers to regulate domestic colonial affairs in no way impaired their right to do so. Moreover, the British had no reason to believe that the colonists saw these issues any differently. The British believed that they had the discretion to take back or otherwise alter these privileges. Finally, in the British view, sovereignty could not be divided. Thus, if anyone had it, they did, not the colonies.

As historians emphasize, the British and the Americans held mutually inconsistent views about their world – with an explosive potential. As the crisis unfolded, the Americans' constitutional interpretation led them to view the British intervention into their domestic affairs as potential tyranny, whereas the British constitutional interpretation led them to view American resistance to their policies as potential treason. Each side considered itself as acting lawfully and legitimately while the other was wholly in the wrong.

Our model of a self-confirming equilibrium helps explain and elucidate the two sides' contradictory interpretations, including how they led to revolution and war. First, we show that prior to the Seven Years' War (1756–63), the British and the Americans were in a self-confirming equilibrium. Although the two sides had very different ideas about the structure of the empire and about British authority, prior to the war, the British had no incentive to exercise their powers. Thus, the occasion that would generate evidence of the conflicting interpretations never arose. Nothing about the two views came into conflict, so neither side obtained evidence from the other to indicate that its conjectures clashed with the other's views or possible actions.

The Seven Years' War produced a dramatic shock to the environment. Prior to the war, the pervasive French threat in part underpinned trans-Atlantic cooperation.<sup>4</sup> Both sides needed each other and were willing to bear costs to maintain the mutually beneficial relationship. The removal of the French threat at the end of the Seven Years' War altered the basic structure of the empire, including both sides' willingness to cooperate. The British gained a clear interest in regulating domestic colonial affairs, in part as a means of controlling a much larger empire that now included French Canada and the large territories containing Indians hostile to the American colonists.

The new configuration of interests following the Seven Years' War brought to the fore the contradictory views over ideals and the mechanisms protecting American rights. What was previously off-the-path behavior for the British – intervening in the American domestic affairs – became on-the-path, much to the surprise of the Americans. To the British, this exercise of authority did not constitute defection from cooperation, and they had no reason to expect that the Americans would resist. The Americans, however, did interpret intervention as defection. In reaction, and regardless of the magnitude of the defection, the Americans therefore initiated a trigger strategy. What seemed to the British to be a relatively mild exercise of their rights was perceived by the Americans as a violation of constitutional rights guaranteed to them by previous commitments. To British eyes, alternatively, American resistance appeared as a bid for independence that sought to explore the removal of the French threat. American resistance thus appeared as a unilateral defection from cooperation, and the British also reacted by exercising a trigger strategy. Put simply, each side viewed the other as the first to defect from cooperation.

Paralleling the historians' arguments, the model reveals how each side had an inadequate mental model of the other. Yet, for over a century neither side obtained any evidence to the contrary – no evidence emerged to cause either to believe that its views were inadequate and inaccurate. By altering both sides' incentives, the Seven Years' War first generated a surprise to each, ultimately resulting in an escalating conflict, revolution, and war.

We develop our argument as follows. Section 2 provides the background on the revolution, including an account of the historical literature, the role of ideas, and a short narrative of events. Section 3 describes the self-confirming equilibrium framework. Section 4 applies the framework to the American Revolution. The heart of the article, this section shows how

---

4. It is worth acknowledging that at times during this period, there was also a distinct threat from the Indian nations to both British and particularly American interests (see e.g., Morgan, 1956).



the American and British views could differ yet not cause conflict during the century prior to the Seven Years' War, only to explode into conflict after the war. Our conclusions follow.

## 2. Constitutional Ideals During the American Revolution

Modern historians (Bailyn, 1967; Wood, 1969; Greene, 1995b; Morgan, 1995; Reid, 1995; Rakove, 1996) emphasize the centrality of ideas and ideology in the development of revolutionary sentiment in America. The colonists devoted considerable effort to theorizing about liberty, the constitution, sovereignty, and the nature of the empire. Indeed, the dominant view among historians is that the revolution took place because of differences in ideals. In the words of one prominent historian:

The Revolution becomes comprehensible only when the mental framework, the Whig worldview into which the Americans fitted the events of the 1760s and 1770s, is known. 'It is the development of this view to the point of overwhelming persuasiveness to the majority of American leaders and the meaning this view gave to the events of the time, and not simply an accumulation of grievances,' writes Bailyn, 'that explains the origins of the American Revolution'.

... If the origin of the American Revolution lay not in the usual passions and interests of men, wherein did it lay? ... Never before in history had a people achieved a 'revolution by reasoning' alone. The Revolution was thus essentially intellectual and declaratory ... (Wood, 1966:162-9)

The debates between the British and colonists over the ideals of constitutionalism, sovereignty, liberty were not simply debates about abstract ideals, however. Before drawing the link between these ideals and everyday politics and economics, we study the ideas' content.

### *Contrasting Views of the Constitution*

To a large extent, both sides shared the same political ideals and traditions. Both accepted that abuse of power by government was a crucial problem, and that it could best be controlled by rule of law. In what follows, we will discuss the ideas that came to dominate in the American colonies and in Britain. This treatment ignores considerable differences of opinion on each side of the Atlantic. In our companion article, we explain how these ideas came to dominate in the colonies (Rakove et al., 1999).

Underlying the broad agreements on ideals, were deep disagreements over the exact nature of the institutionalization of rule of law, particularly as it applied to the empire. For example, the British and Americans disagreed about the substantive American rights protected by the British constitution and the institutions established to protect those rights. Over time, the two sides also came to disagree about each other's commitment to those ideals.

Put crudely, the colonists worried that the new policies proposed by the British after the Seven Years' War signaled a change in the constitution of the empire. They feared that constitutional understandings that could be easily changed provided little protection of their rights. In light of these changes, colonists sought assurances that the British would respect their liberty. At the same time, the British worried that resistance to the new policies signaled a change in colonial attitudes toward the empire. They worried that claims about fundamental rights could be expanded easily, and thus provided a weak foundation for membership in the empire. In light of these changes, imperial officials wanted assurances that any concessions would not lead to American independence.

### **The American View**

The colonists derived their view of the imperial constitution from their experience. From the start, the British colonies were unlike those of Spain or Portugal. The British relied on large-scale settlement by adventurers, lured by the promise of a better life. Since that better life was not a sure thing – life was unusually harsh for the early settlers – the King had to offer various concessions to lure potential colonists. Among these were grants of local political power, including that of a local legislature.<sup>5</sup>

In the century prior to the revolutionary crisis, the *de facto* power of the colonial assemblies continued to grow. Although the King appointed most colonial executives, he was dependent in important ways on the colonial legislatures. This dependence reflected economic necessity: given the costliness of trans-Atlantic communication, such decentralization was the only viable alternative to a metropolitan control that would simply have been too expensive to be effective.

At the same time that practice was evolving toward what Greene (1991) calls 'negotiated authority', that is meaningful decentralization, law was continuing to celebrate the form of centralized power. Thus, during the 17th and early 18th century, various practices provided the King with authority over the colonies. The crown continued to appoint most colonial governors, and the acts of all colonial assemblies were subject to review by the Privy Council (Smith, 1965). In practice, however, that authority was largely a dead letter. For example, governors were routinely instructed to seek the aid of the assemblies. The decline of imperial participation culminated in the Walpole ministry's adoption of what Burke later called a policy of 'wise and salutary neglect' during the 1730s. As a result, by 1750, the institu-

---

5. The absence of valuable treasure – a major factor determining the structure of the Spanish empire – made these concessions less costly for the English crown than for the Spanish crown (see North et al., 1999).

tions charged with overseeing the details of colonial policy, such as the Board of Trade, had 'atrophied' (Greene, 1991: 97). In short, with respect to their internal affairs, the colonies were self-governing.<sup>6</sup>

Of the political practices that had grown up to govern the empire, the colonial assemblies were at the heart of many Americans' concerns over liberty. Although Americans disagreed about any number of issues, they commonly agreed that their legislatures were central to protecting these ideals. Members of self-sufficient religious communities, such as those in rural Massachusetts, might be indifferent to British imperial trade regulation, but they believed that colonial legislatures were central to protecting their religious freedoms that distinguished America from Britain. Similarly, many merchants might care far more about British imperial trade policy than any other issue, but they depended on their legislature to maintain low and stable taxes and to define and protect their property rights and wealth. Whatever Americans valued as their liberty, the colonial legislature was central to protecting it.<sup>7</sup>

For the colonists, the fact that they had long exercised authority through their local legislatures established a constitutional claim. They argued that the British constitution was like the common law, a set of rules built on practice and precedent. Drawing on earlier Whig traditions, the colonists argued that custom and precedent were effective ways to restrain arbitrary rule. In arguing for custom, they were not, as many modern scholars argue, indulging in a nostalgic attachment to the past. Rather, they saw custom as a commitment device, since it, unlike statutes, could not be changed by any small group, such as the Crown or Houses of Parliament.

From this perspective, the historic principles governing the empire, at least as interpreted by the Americans, gave them a degree of protection – indeed, of credible commitment – against the imperial government. The notion of divided sovereignty reflected, in effect, a view that the empire had a federal structure: Britain controlled truly empire-wide issues, such as trade and security, while Americans controlled local matters, such as economic and social organization, through their legislatures. As long as the Americans and the British held to this structure, the principal American values and freedoms – including their livelihoods, their property and their religious practice were protected from arbitrary action by the British crown or Parliament.

---

6. In fact, as Weingast (1995) argues, the division of authority in the empire matched a type of federalism called market-preserving federalism.

7. Not all Americans agreed that legislatures were central to the protection of freedom. Many minorities, anticipating the argument of Federalist #10, feared that legislatures would become tools by which majority factions would oppress minorities. For them, the British Empire provided a 'controlling common umpire', without which the colonies would become 'a field of blood, a scene of terror and desolation' (Potter, 1983: 35, 143).

From the colonial perspective, a constitution founded on precedent embodied a 'bright line credible commitment' mechanism in which there are clearly understood actions which constitute defections from cooperation, and prescribed punishments for defections (see e.g., de Figueiredo and Weingast, 2005). The federal structure of the empire had the advantage of making deviations easy to observe (hence the 'bright line') and thus easy for both sides to police. The division of authority inherent in the federal system governing the British empire served as a bright line, a rule which was so clear that virtually any citizen could tell when it had been breached. As coordinated reactions were the only hope for the American colonists to police the British, the bright line mechanism facilitated a coordinated reaction to breaches.

### **The British View**

Although the British based their views of the imperial constitution on the same practices as the colonists, they drew very different lessons from those practices. The British regarded their domestic institutions of governance, especially the King in Parliament, as the chief bulwark against tyranny. Over the 18th century, this commitment led them closer to articulating the modern doctrine of parliamentary sovereignty, under which any act of Parliament could supersede any existing constitutional understanding (Blackstone, 1765: 156).

From the British perspective, precedent and practice had very different constitutional standing than they did for the Americans. To the British, the governance structures of the empire, while of great value, did not have constitutional status. These structures were a matter of policy dictated by expediency and efficiency. As such, Parliament could alter the empire's governance structure whenever it wanted. This view implied that the British believed that the powers exercised by colonial assemblies were privileges granted to the Americans, not constitutional rights restricting parliamentary authority. Finally, the British believed that sovereignty was indivisible, implying that, through its Parliament, Great Britain was sovereign over all aspects of the empire.

### *The Constitution and the Revolutionary Crisis*

Until the end of the Seven Years' War, American and British differences led to few lasting problems. As Greene (1995a) documents, many disputes over authority arose during this period, as colonists clashed with imperial officials on a wide range of issues. Before 1763, however, the metropole and colonies had managed to resolve these disputes peacefully.

Following the long series of wars from the late 1730s that ended with the Seven Years' War, British imperial policy underwent significant change, reflecting the changing demands of empire. Although the governance of

the empire, American in particular, changed after 1763, the British believed the change fully within their constitutional authority. They had fought an expensive war in part to protect the colonies; the demands of a huge new empire required significant reform. From the British perspective, the American demands and arguments seemed flagrant violations of the imperial system. Despite the Americans' claims of affirming the status quo, their demands appeared to the British as a call for revolutionary separation from the empire (Tucker and Hendrickson, 1982).

### *A Short Narrative of Events*

The British and the colonists reacted to the shock of the end of the Seven Years' War against this background (Christie, 1966; Tucker and Hendrickson, 1982). The end of the War and the removal of the French from North America had important implications for both the British and the colonists. First, it decisively ended the French threat to mainland settlement. Second, victory led to creation of a much larger, more secure, more diverse – and more deeply indebted – empire. Third, the first two implied major changes in the interests of both Britain and America. Self-interest no longer underpinned a harmony of interests, in which a harm to one also harmed the other.

The Seven Years' War had several implications for American behavior. First, in the absence of the French threat, Americans would be less likely to suffer burdens, because there would no concomitant gains. The same held for the British: the now much larger empire implied that the original American colonies no longer loomed so large in the British imperial calculus. Standard theories of bargaining among self-interested parties suggests that the Americans would be less conciliatory to the British than before the war. Second, additional economic changes occurring in America: by mid-century, for the first time, America had real wealth to expropriate (McCusker and Menard, 1985). This, too, had an implication for both the British and the Americans. For the British, it implied that for the first time there was something to extract from the Americans; for the Americans, it implied that they had more to fear from British exploitation.

At the beginning of the crisis, the radicals emerged to argue that the British were malevolent and out to destroy American liberty. Yet these arguments failed to sway many. These seemingly wild claims simply failed to match experience. Over time, however, the British behavior came to take on a different cast. A series of acts seemed clearly to signal a change in British policy with respect to American autonomy, the constitution, the empire, and the role of the British in regulating domestic colonial affairs. This began with the Sugar Act of 1764, explicitly designed to raise revenue, and the Stamp Act, which would tax Americans directly. Although the magnitude of

these taxes was relatively trivial, American radicals argued that they reflected a sea-change in British authority and intention.

As is well-known, colonial resistance forced the British to reconsider their policy, revoking these policies in short order. In the face of the Stamp Act's failure, however, the British passed another act even as it removed this tax. The Declaratory Act of 1766 held that Parliament was unlimited and unlimitable (Greene, 1995a: 85), arguing that the removal of the tax was not intended to signal that the British agreed with the Americans that they lacked the power to impose such taxes, just the opposite.

The next step in the cycle came with the enactment of the Townshend duties in 1767. Exploiting a distinction some colonists (including Franklin) had drawn between unacceptable internal taxes and legitimate external ones, such as duties, this measure eventually led the colonists to implement a new boycott of British imports. John Dickinson's influential Letters from a Farmer in Pennsylvania helped clarify the colonial position. Any revenue measure, internal or external, imposed on the colonists without their consent violated the existing constitutional understanding.

The next major policy initiative came in 1773, with the passage of the Tea Act. To ease the woes of the British East India Company, Parliament lowered the duty on tea, while also beefing up the penalties for smuggling. It did so in the hope that the tax would be so low that Americans would set aside their principles and pay the duty.

Once again, the colonists protested the constitutionality of the Act. This protest culminated in the Boston Tea Party, during which the colonists destroyed tea in defence of their constitutional claims. Once again, the British reacted to this resistance by stepping up their enforcement with passage of the Coercive Acts (Greene, 1995a: 23). The initial bill closed the Boston harbor until it paid for tea and until duties were collected. Parliament adopted three further measures. The Massachusetts Government Act annulled the charter of the Massachusetts colony; the Administration of Justice Act provided that imperial officials would be tried outside of the colony; and the Quartering Act allowed the imperial officials to seize property to support troops if the colonial assembly failed to allocate the necessary funds.

Finally, the Quebec Act expanded new holdings in Canada into the territory claimed by colonies, attempting to prohibit American colonial expansion into trans-Appalachia. It also legislated religious tolerance for Catholics. Both provisions raised constitutional issues for the Americans. The prohibition on immigration unilaterally altered colonial charters by affecting their boundaries. More ominously, given that many colonies were founded on sustaining religious freedom, Parliament's regulation of religion threatened the American separation of church and state. If the British had

the constitutional authority to regulate religion in Quebec, it possessed the authority to do so in the American colonies.

The differing American and British conceptions of sovereignty and the constitution guided both sides in their interpretation of their own and their opponent's actions. The colonists' view about the long history of divided sovereign authority within the empire meshed with their view of the constitution, based on precedent and long-established practice. From the colonial perspective, by breaking with precedent, the British attempts at taxation, quartering troops, suspending colonial legislatures, altering colonial charters, and regulating religious practice seemed flagrant violations of the constitution and American liberty. In the mid-1760s, the radicals' predictions that the British would undermine the centrality of the colonial assemblies seemed outlandish. By the mid-1770s, the predictions had come true. The weight of the British actions thus altered the moderates' perception of British malevolence, pushing them into the radicals' camp (Rakove et al., 1999). By the end of 1774, it was hard for the moderates to sustain the view that the British were benign.

### *Implications of These Events*

From the beginning of the crisis, radicals sought radical action. But this faction was too small to be politically dominant. Those loyal to Britain occupied the other end of the political spectrum. In between these two factions stood the moderates, a group of prudent and sober individuals who understood the risks of revolution and war (Rakove, 1979). The moderates were pivotal politically: if they remained loyal, no effective challenge to the British could be mounted. But if they joined the radicals, revolution would gain sufficient support.

The events described here had a cumulative impact on moderate Americans. In our companion article, we use a simple Bayesian model to show how the various events just noted would have altered the moderates' beliefs (Rakove et al., 1999). Whereas the moderates were inclined in the mid-1760s to believe the radicals imprudent, by the mid-1770s, many had changed their minds. Regardless of whether the British were actually intent on expropriating colonial wealth and closely regulating colonial society, the British had become a malevolent presence willing to suspend or alter the heart of the political and constitutional mechanisms that Americans believed protected their liberties and their way of life. The British willingness to regulate a range of domestic affairs – potentially including religion – and their willingness to abrogate the foundation of American liberty – the colonial assemblies and their role in creating American law – represented a direct threat to the American way of life. The empire's division of powers served as a bright-line commitment mechanism.



Understanding this principle helps explain an important component of the colonists' fears generated by the British actions and newly articulated constitutional principles after 1763. The British view that Parliament was unlimited and unlimitable threw aside the mechanisms providing credible commitment for American liberty. Moreover, from the colonial perspective, the British theory of Parliamentary sovereignty offered *no* credible substitute, not even a poorer one. This omission potentially left Americans without credible protection from opportunistic action by the metropole. Because today's parliamentary majority could not bind tomorrow's, anything enshrined in today's legislation could be altered tomorrow. As the radicals observed early in the controversy, if the British could exercise these powers in mild and relatively innocuous ways, they could do so in ways that expropriated Americans or which began to regulate social fundamentals, such as property and religion.

Further, the absence of a credible compromise from the British – indeed, the absence of any attempt to court the American moderates away from the American radicals (Rakove, 1979) – provided further evidence in favor of the radicals' thesis. If the British were not malevolent, why did they not offer a reasonable compromise?

The role of ideas proved central to the moderates concluding that they needed to side with the radicals. The ideals of constitution, sovereignty, federalism in the empire, and above all, American liberty, gave Americans the framework within which to interpret British action. American theorists explained how these mechanisms provided credible protection in the empire and how the new British views directly threatened Americans. American moderates therefore came to see great danger in the British dismantling of the former mechanisms of credible commitment.

Still, a number of puzzles remain. How could the British and the Americans sustain these differences over the previous century? In particular, how could they cooperate for so long in the face of these differences? And what changed following the Seven Years' War, underpinning the explosive conflict?

### 3. The Self-confirming Equilibrium Framework

This section contains the technical preliminaries that define a self-confirming equilibrium and its relation to the more traditional Nash equilibrium framework.<sup>8</sup> One purpose of the approach is to allow the player's mental models of their strategic situation – as modeled by their beliefs about the extensive form

---

8. Fudenberg and Kreps (1995), Fudenberg and Levine (1993, 1998), Greenberg (1994), Kalai and Lehrer (1993) explore this and related concepts. Sargent (1999) applies self-confirming equilibria to macro-economic phenomena.



– and about the intentions of others – as modeled by their strategies – to differ. These variations allow us to model players with cognitive limits in the sense identified by North (1999a) – they have incorrect and inadequate understandings of their environment and about other players’ intentions. The equilibrium aspect of the framework provides the discipline that prevents these beliefs from being arbitrary, wildly incorrect, or mutually incompatible.

Several participants in this literature explain the motivation for this approach. Fudenberg and Levine (1993) weaken the Nash equilibrium requirement that each player holds correct beliefs about their opponent’s strategies. They require ‘only that players’ beliefs are correct along the equilibrium path of play. Thus each player may have incorrect beliefs about how his opponents would play in contingencies that do not arise when play follows the equilibrium, and moreover the beliefs of different players may be wrong in different ways’ (p. 523). In discussing these and related notions, Greenberg (1994) suggests that ‘these notions are motivated by the fact that “off equilibrium choices” are not observed, and hence the requirement of “commonality of beliefs” cannot be justified’ (p. 4).

In what follows, we adopt a variant of this approach which most closely resembles the approach suggested by Greenberg et al. (2004; see also Greenberg, 1994). In particular, we allow players to have differing beliefs about the extensive form, though in equilibrium, these differences occur only off the path. In this case, players may have limited knowledge about one another’s capabilities. In other words, players may be unaware of the actions other players could take. As an example, given two distinct players  $i$  and  $j$ ,  $i$  may not know the full set of actions that  $j$  may take. In equilibrium,  $j$ ’s strategy will depend on the full set of actions she may avail herself of. On the other hand,  $i$  may never learn of  $j$ ’s feasible move if that move is never used or observed. In this sense, unlike in the standard Nash equilibrium framework, a strategy profile for a player may not be contingent on the full set of actions opponents can take.

The framework begins with an extensive form game,  $M$ , and a set of  $N$  players. Each player  $i$  has a set of beliefs,  $\mu_i$ , about the structure of the extensive form game. These beliefs may differ from one another and from the actual extensive form, though as we will see, in equilibrium, they cannot differ ‘too much’. Given a history of the game up to move  $k$ ,  $h_k$ , a **strategy**,  $s_i$ , for player  $i$  is a rule that specifies a proposal for any move at which  $i$  may make one and is a function of  $h_k$ . Each player also has a set of beliefs,  $\sigma_{-i}$ , about the strategies being played by the other players on  $\mu_i$ .

A **self-confirming equilibrium** is a set of strategies,  $s^* = (s_1^*, \dots, s_n^*)$  and beliefs  $(\mu, \sigma) = (\mu_1\sigma_1, \dots, \mu_n\sigma_n)$  with three characteristics:

- (a)  $U^i(s_i^*/\sigma_{-i}) \geq U^i(s_i/\sigma_{-i})$  for  $s_i \neq s_i^* \forall i$ ; further, this property holds for all proper subgames of  $\mu_i$ ;
- (b)  $\mu_i$  must be identical to  $M$  along the equilibrium path; and
- (c)  $\sigma_{-i} = s_{-i}^*$  along the equilibrium path of  $\mu_i$ .

Condition (a) requires that player  $i$ 's strategy maximizes her payoffs, given her beliefs about both the extensive form,  $\mu_i$ , and the strategies being played by others,  $\sigma_{-i}$ . This condition also requires that her strategy is subgame perfect on  $\mu_i$ . Conditions (b) and (c) require that no events along the path disconfirm  $i$ 's beliefs about the structure of the extensive form or about her opponents' strategies. Notice that condition (b) implies that  $\mu_i = \mu_j$  along the equilibrium path, though they may differ off the equilibrium path.

Pedagogically, one can think of a game with a self-confirming equilibrium as  $n + 1$  games: the game that each of the  $N$  players believes she's playing ( $\mu_i$ ) plus the actual game they are playing. A self-confirming equilibrium in this context can be broken down into several parts. Consider player  $i$  and the game she believes she is playing,  $\mu_i$ . First, given the strategies  $i$  believes the other players are playing ( $\sigma_{-i}$ ),  $i$ 's strategy is an optimal response. Moreover, the strategy that  $i$  believes each of the others is playing must also be an optimal response for that player on  $\mu_i$ . Second, this set of strategies ( $s_i, \sigma_{-i}$ ) constitute a subgame perfect equilibrium on  $\mu_i$ . Third, the set of strategies each player actually plays ( $s_1^*, \dots, s_1^*$ ) must constitute a self-confirming equilibrium on the actual game they are playing ( $M$ ). That is,  $s_i$  is an optimal response for  $i$  on  $\mu_i$ ; along the equilibrium path,  $\mu_i = \mu_j \forall i, j$ ; and  $\sigma_{-i} = s_{-i}^*$  along the equilibrium path of  $M$ .

To understand a bit more about self-confirming equilibria, it is useful to compare them with Nash equilibria. Self-confirming equilibria represent a generalization of Nash equilibria in the sense that all Nash equilibria are self-confirming equilibria. The latter differ from Nash equilibria in at least two respects.<sup>9</sup> First, the Nash framework assumes that the game's structure is common knowledge. Second, this approach also assumes that the players know the other players' strategies or are aware of their ignorance. In technical terms, the common knowledge feature of Nash implies that  $\forall i, \mu_i = M$ . The assumption about strategies requires that,  $\forall i, \sigma_{-i} = s_{-i}^*$  over all  $M$  not just along the equilibrium path.

In contrast to the Nash framework, nothing in the self-confirming equilibrium framework forces complete agreement among the players' beliefs about the structure of the extensive form and about the strategies played

9. Fudenberg and Levine (1998) note other differences, such as different learning properties traced through asymptotic techniques.

by others. Self-confirming equilibria allow player  $i$  to play a strategy different from that which player  $j$  believes  $i$  is playing, where the differences occur off the equilibrium path. Because each player observes only play along the equilibrium path, neither becomes aware of their inaccurate understanding of the game. Yet these potentially contradictory beliefs are an equilibrium because no player observes anything to contradict their beliefs about  $\mu_i$  or  $\sigma_{-i}$ . As Greenberg (1994) suggests, the players may disagree about why equilibrium arises and, further, they may not know they disagree. Player  $i$ , for example, may choose move  $A$  over move  $B$  because of her conjectures about player  $j$ 's response. But player  $j$ 's response is never observed and may in fact be different than  $i$  anticipates.

An important feature of the definition of self-confirming equilibrium constrains the beliefs about the extensive form game,  $\mu_i$ , and about the strategies of others  $\sigma_{-i}$ . These beliefs cannot be so wildly off that an individual is called to make a move she did not believe existed or that a player observes another player taking an action that the first believed infeasible (i.e., not part of  $\mu_i$ ). Along the equilibrium path, nothing about the play disconfirms player  $i$ 's beliefs about the extensive form or about the strategies played by the other players.

### *A Model of Surprises*

An important aspect of this framework is that it allows surprises. Players in a self-confirming equilibrium may have inaccurate beliefs about the extensive form,  $\mu_i$ , or the strategies being played by their opponents,  $\sigma_{-i}$ . As long as the system is stable, these inaccurate beliefs may not be revealed.

If the system undergoes an exogenous shock, however, then these inaccurate beliefs may come to the fore. We model this shock by allowing nature to transform the game some time after the initial self-confirming equilibrium has been in force – presumably, long after. An example is a shock that alters the payoffs for some or all of the players. The change in payoffs may then cause one or more players to alter their behavior, that is, to change their action choice in a particular move of the game. In terms of the model, something that was previously off the path becomes on the path.

Exogenous shocks of this type imply that, if one or more of the other players have inaccurate beliefs about one or more of the other players' off-the-path play, they may mis-predict how that player reacts to the shock. Because the self-confirming equilibrium framework allows players to be misinformed about one another's types, mis-predictions after an exogenous shock can lead to surprising behavior. As in the Nash framework, players in the self-confirming equilibrium framework may not know precisely another player's preferences. In contrast to the Nash equilibrium framework,

however, players in the self-confirming equilibrium framework may be completely ignorant of even the probability distribution from which these preferences are drawn and, even further, may not be aware of their ignorance.

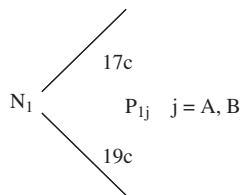
#### 4. A Self-confirming Equilibrium Model of the American Revolution

In this section, we apply the self-confirming equilibrium framework to the American Revolution. This application constitutes an ‘analytic narrative’ (Bates et al., 1998), that is, it uses a model to interpret a single case.

We begin by describing a two-person game facing the British and the Americans. The basic interaction takes place in what we call the ‘British empire stage game’. The stage game is embedded in a context where nature has two moves, one at the beginning of the game, and one prior to the play of the stage game in a distinguished time period,  $t^*$ .

Prior to the first play of the game, Nature has the first move in which it must choose a type for the British. Although the British know that Nature makes this choice and learn its outcome, the Americans are unaware that this choice is being made. We represent this choice as a move by Nature ( $N_1$ ) in Figure 1. Nature’s choice of a British type results in the constitutional view held by the British, which may be either of two types. The first type is that the British retain throughout the entire period the status quo view of the constitutional setup that prevailed following the Glorious Revolution in the late 17th and early 18th centuries (labeled, 17c). The second type is that the British constitutional view slowly evolves between the beginning of the game and period  $t^*$  into a newer constitutional view that rose to maturity in the 19th century (labeled 19c). The consequence of these views are described later; as suggested earlier, they have different implications for the understanding of British powers and actions.

The probability that Nature chooses the 17c view is  $p_{1j}$ , where  $j = A, B$  for the Americans and the British, respectively. The British accurately assess the likelihood of Nature’s choice, so that  $p_{1B} = p_1$ , where  $0 < p_1 < 1$ .

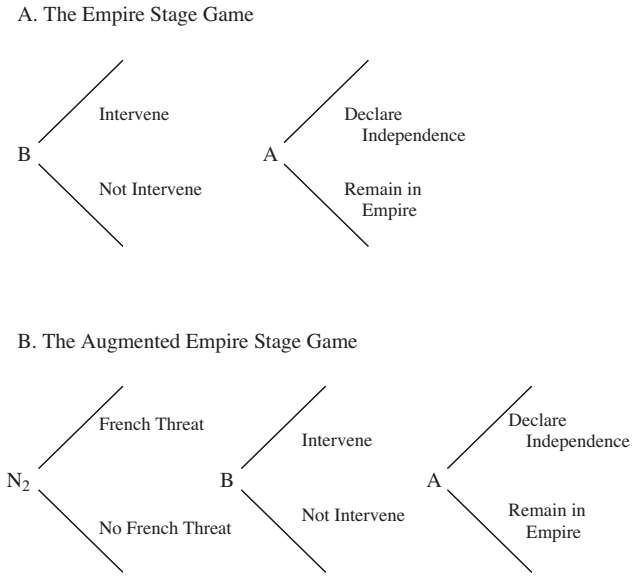


**Figure 1.** Nature’s Choice of a British Type

The Americans, however, are unaware that this choice is being made. They therefore assume that the probability is identically equal to zero; that is,  $p_{1A} = 0$ . This first move by Nature represents the first of two principal departures of this game from the common-knowledge structure of Nash equilibrium games. By assumption, the Americans are unaware that the British type is uncertain and hence unaware that the British payoffs are uncertain. Put simply, their mental model, as expressed in their understanding of the extensive form and hence of the range of possible types they may face, inaccurately reflects their strategic environment. Similarly, this departure from common knowledge itself is not known to the British: they are ignorant of the American's ignorance of the move by Nature at the beginning of the game.

Turning to the empire stage game, the British move first (see Figure 2a). They must decide whether or not to intervene in the domestic affairs of the American colonies. As we will see, there may be multiple interpretations of the meaning of this intervention, but for now we suppress them. The Americans move next, deciding whether to remain in the empire or whether to declare independence.

The stage game illustrated in Figure 2a is played in every period except a distinguished period,  $t^*$ . In that period, Nature again moves, this time just prior to the play the stage game (Figure 2b). Nature's ( $N_2$ ) choice in this



**Figure 2.** The Empire Stage Game



The two (independent) choices by nature determine four separate cases, whether the British constitutional type is that of the seventeenth century (17c) or the nineteenth century (19c) and whether there is a French threat (Fr Th) or not ( $\phi$ )

**Case 1A: 17c, and French Threat**

		British	
		Not Intervene	Intervene
Americans	Remain	5, 5	-3, 7
	Declare Independence	7, -3	0, 0

**Case 1B: 17c, and No French Threat**

		British	
		Not Intervene	Intervene
Americans	Remain	2, 2	-1, 2.5
	Declare Independence	2.5, -1	.5, .5

**Case 2A: 19c, and French Threat**

		British	
		Not Intervene	Intervene
Americans	Remain	5, 5	-3, 7
	Declare Independence	7, -3	0, 0

**Case 2B: 19c, and No French Threat**

		British	
		Not Intervene	Intervene
Americans	Remain	2, 2	-1, 6
	Declare Independence	2.5, -1	.5, 3

**Figure 4.** Payoffs from the Empire Stage Game

As can be seen, the empire stage game is represented by a prisoner's dilemma in three of the four cases (all but case 2B). In all four cases, the British have the choice of intervening or not in the Americans' domestic affairs, and the Americans have the choice of remaining in the empire or declaring independence. In the presence of the French threat, both sides are better off cooperating than not cooperating, receiving a payoff of 5 each play (0 if they do not cooperate). Still, each player can take advantage of the other in the short run. If the British choose unilateral defection – that is, to intervene while the Americans cooperate – the British get 7 while the Americans get  $-3$ . If the Americans unilaterally defect – that is, to declare independence – and the British cooperate, the Americans get 7 while the British receive  $-3$ . Because the British hold the 17th-century view of the constitution in this case, they, like the Americans, view intervening in the domestic affairs of the American colonies as a violation of the constitution. Because  $N_2$  chooses to maintain the French threat in case 1A, the payoffs to the stage game are constant in all periods (before and after period  $t^*$ ).

In case 1B, the payoffs are the same as case 1A for time periods 1 through  $t^* - 1$ . In period  $t^*$ ,  $N_2$  chooses that the French threat disappears, so that the payoffs switch to those listed under case 1B in Figure 4. The stage game under case 1B also has the structure of a prisoner's dilemma, yet the payoffs have changed. In particular, the value of cooperation between the British and the Americans has declined (though is still positive). In this case, cooperation is worth 2 to each, non-cooperation is worth 0.5. Put simply, in the absence of the French threat, the Americans and the British need each other less. Unilateral defection gives the defector 2.5, while the sucker a payoff of  $-1$ .

In case 2A, the payoffs are identical to those in case 1A. Although  $N_1$  chooses the British type as 19c, this does not affect the game, as  $N_2$  chooses to retain the presence of the French threat.

The main difference in the cases occurs in case 2B, when  $N_1$  chooses the British constitutional type as 19c and  $N_2$  chooses to remove the French threat. The strategic situation here differs considerably from the other three cases. As in case 1B, the move labeled cooperation is worth 2 to both players. Unilateral defection by the Americans has the same payoffs as under case 1B, 2.5 to the Americans and  $-1$  to the British. The principal changes in payoffs occur when the British defect. Unilateral British defection represents intervention into the American economy. When the British type is 19c, its payoff to this defection is 6, not 2.5 as in case 1B. If both players defect, the Americans receive 0.5 and the British, 3.

Finally, the payoffs in the repeated game are simply the discounted sum of the stage game payoffs. We designate the discount factor by a common parameter  $\delta\epsilon(0, 1)$ .



### Equilibria

This game possesses multiple equilibria, and we do not attempt a full characterization. Mutual defection, for example, is an equilibrium throughout. We are interested in the ability of the players to maintain cooperation, particularly as their environment changes as represented by nature's choice about the French threat. Hence we focus on equilibria in which cooperation can be sustained.

Following the definition of self-confirming equilibrium, we write the strategies contingent on those choices by nature of which the players are aware. Thus, the British strategy will depend on both choices by Nature. In contrast, the Americans are unaware of Nature's first choice, so their strategy will not be contingent on  $N_1$ 's choice. Thus, the game the Americans believe they are playing differs from the game the British believe they are playing, that is,  $\mu_A \neq \mu_B$ . Our strategy in solving this game is to suggest that the Americans and the British were in a self-confirming equilibrium for generations prior to the distinguished time period,  $t^*$ , when nature removed the French threat. Because the Americans are unaware that Nature chose a British type, they do not accurately predict British behavior under all circumstances following Nature's choice about the French threat at time  $t^*$ . To be a self-confirming equilibrium, however, nothing about the British behavior disconfirms the Americans' beliefs about the British strategy or the structure of the game.

The equilibria are described as follows. First, suppose that  $N_1$  chooses 17c. In this case, for  $\delta$  sufficiently large, cooperation can be sustained, regardless of whether  $N_2$  chooses to continue or remove the French threat. Cooperation is supported by the usual trigger strategies so that each player cooperates in period 1 and, thereafter, cooperates as long as their partner does so in the previous period. If ever a player's partner fails to cooperate, then that player will defect for ever.

Second, suppose that  $N_2$  chooses 19c. In this case, the British strategy is contingent on  $N_2$ 's choice in a different way. In the absence of the French threat, the value of intervening in American domestic affairs is larger for the British. And, for appropriate values of  $\delta$ , given the low payoff from cooperation relative to British defection, cooperation cannot be sustained where it previously could be. Thus, contingent on  $N_2$ 's choice to remove the French threat ( $\phi$ ), the British switch their actions from non-intervention to intervention.<sup>11</sup>

---

11. It is interesting to note that there are three regions of  $\delta$  which determine the classes of equilibria which can be sustained. For very high values, cooperation can be sustained under all of the cases. For sufficiently low values, cooperation cannot be sustained under any of the cases. For intermediate values, however, cooperation can be sustained for all but case 2B.

The Americans, on the other hand, are blithely unaware in this case that the British constitutional view is 19c, not 17c. According to their beliefs,  $\mu_A$ , they are in case 1B. The Americans' strategy is therefore the same as in case 1B; and further, they believe that the British strategy is the same as in case 1B. The Americans believe that the British have an interest in cooperation in the absence of the French threat. Whereas in fact, when the British are type 19c, they do not want to cooperate. Thus, a real surprise emerges in Case 2B at time  $t^*$ , when  $N_2$  chooses to remove the French threat and the British choose to intervene in the American domestic affairs.

This equilibrium encompasses a pair of mutually inconsistent expectations in case 2B where the French threat disappears. The British believe they have the power to intervene and have no reason to believe that the Americans will resist. Their conjecture is wrong, but for generations prior to the removal of the French threat, this part of the game tree was off the path and thus never observed. The Americans were unaware of the uncertainty about the British type and thus that the British constitutional beliefs had changed. As the British never chose to exercise the authority they believed they had, the Americans never obtained evidence of the new British constitutional views. In short, both sides mis-conjecture in equilibrium about the other's behavior.

It is useful to compare our framework to a benchmark case of standard game theory: a Perfect Bayesian Equilibrium (PBE). One variant of our model could be that the Americans do not know the move by  $N_1$ , but instead they assign a very low likelihood to a 19c British type. In this case, if either the probability that the British are of type 19c or the probability of the French threat disappearing was very low, then it could be the case that a Perfect Bayesian Equilibrium could yield cooperation in the periods prior to  $t^*$ . Our framework differs in a number of important ways, however. First, from a substantive point of view, the approach we take is interpretively different: while the Americans would know there is a *possibility*, even if improbable, that the British may be 19c in this PBE framework, the knowledge of the Americans in our framework is fundamentally different since they are not even aware of the possibility. In this sense, surprise is more than simply the occurrence of an improbable event – it is the occurrence of a previously *unknown* event. Second, our framework therefore allows for a more direct model of the interpretation provided by historians concerning *ideas*. In the case we lay out, an idea is each side's understanding of what the game is. In the PBE framework, since the game is common knowledge, this interpretation cannot be accommodated. Third, unlike in a PBE, our framework allows us to explain surprise about the classification of actions. Critical to our interpretation is not only that the Americans are surprised by an improbable event, but that while the British believe they are not defecting in their limited interventions, the Americans do classify British actions as defection from cooperation. Similarly, the British perceive American

resistance as a bid for independence. Thus, the surprise is two-fold, by both British and Americans, which would not be accommodated in the standard PBE formulation. Finally, because the move by  $N_1$  is necessarily on the equilibrium path of play, the PBE framework puts restrictions on how large  $p_1$  can be. If it is too large, for example, then cooperation could not be sustained. In a variant of our model, however, in which the British know the Americans' ignorance, a large  $p_1$  would not necessarily eliminate the chances for cooperation prior to a shift in the French threat.

### *Implications for the American Revolution*

Our application of the model to the case is as follows. Nature chose 19c as the constitutional type for the British. Although the Americans remained ignorant of this, the British constitutional view evolved from that in the 17th century at the beginning of the game to the beginnings of the 19th century view by period  $t^*$ , representing the defeat of the French at the end of the Seven Years' War in 1763. For a century, the Americans and the British were in a self-confirming equilibrium equilibria where the Americans misunderstood their strategic situation. As long as the French threat remained, however, their misunderstanding did not matter. Both the Americans and the British had a sufficiently strong interest in maintaining cooperation. Thus, by the end of these 100 years, the British had evolved the new 19c constitutional view and thus believed they had the power to intervene in the American domestic affairs. As long as the French threat remained, however, they chose not to do so.

The Seven Years' War fundamentally altered British interests in several ways. The demise of the French lowered both sides' willingness to bear costs to maintain their relationship. Simply put, they needed each other less. Moreover, after the war, the British had a multifaceted empire, no longer dominated by the Americans. The British had to manage French Canada and the Indians in trans-Appalachia. Rules designed by the British to manage the whole might not be optimal for the participants in any of the parts. Finally, the British ended the war with a huge debt and sought to have the Americans pay what they believed was a modest share of this cost.

Following the demise of the French threat, therefore, the British switched choices, seeking to intervene in the colonial domestic economy, albeit in a limited way. Because the Americans did not know the British type, the Americans interpreted this action as a violation of the constitution and thus as a defection from mutual cooperation. The Americans then exercised the contingent trigger in their equilibrium strategy and declared independence. The British, however, did not believe that they had defected from

cooperation.<sup>12</sup> Moreover, because they had no reason to believe that the Americans misunderstood this, they believed that the Americans would pay the small tax to finance their defense.

The British therefore believed, in (a self-confirming) equilibrium that the Americans would acquiesce in their intervention. When instead the Americans resisted, the British interpreted the Americans' resistance as defection from the empire. They reacted, in turn, with harsher measures designed to punish and deter American resistance. The British mis-predicted the American reaction however. Instead of deterring American resistance, British actions strengthened it. The consequence is that both players switched to mutual punishment, and war ensued.

The key to the model is that there are two patterns of behavior, depending on the combination of nature's choices. When  $N_1$  chooses the new type (19c) for the British, the Americans are not aware of it. As long as the French threat remains, however, the British never act on the changes inherent in their type. In this self-confirming equilibrium world, the Americans and the British held contradictory theories about the world. The two sides' beliefs about the extensive form,  $\mu_A$  and  $\mu_B$ , differed from one another. Further, both sides have inaccurate beliefs about the other's strategy after time  $t^*$ . The Americans' mistakenly believe that the British have an interest in continuing to cooperate despite the absence of the French threat. The British mistakenly believe that the Americans will accept their intervention in the form of modest taxes. And yet, as long as the French threat remained, neither side acted in a way that disconfirmed the other's beliefs.

Only when the French threat disappears does conflict emerge and thus the two sides' different constitutional interpretations become relevant. Because the Americans were unaware of the changes in the British constitutional view, they did not correctly anticipate British behavior following the removal of the French threat. The British view and action came as a surprise – indeed, a rude shock – forcing the Americans to radical action.

One may ask whether the above behavior is really an equilibrium – after all, both the Americans and the British are 'surprised' in the play after period  $t^*$ , the defeat of the French. Clearly both sides learn at this time that something in their worldview is false. In a technical sense, the strategies defined before period  $t^*$  are not an equilibrium since they cannot survive the

---

12. In this discussion, we apply the initial move by Nature as determining the understanding by the British of what constitutes a defection. Because the British believed they had the prerogative to intervene in colonial affairs, as long as their interventions were necessary, moderated, and not abusive, they did not interpret limited intervention as violating the social contract they believed existed with the colonies. On the other hand, because the Americans viewed any intervention as a crossing of the 'bright line', the limited intervention by the British was sufficient to constitute a defection.

data observed after  $t^*$ . In a practical sense, however, we believe that this application makes sense. The period  $t^*$  takes place over 100 years after the American colonies were established and over 75 years after the British and the French began fighting in 1689. This represents three generations of mutual coexistence in which nothing either side did disconfirmed the others' beliefs. The behavior modeled by the empire stage game in case 2A prior time  $t^*$  thus occurred over several generations. As our model suggests, this behavior is stable, and reflects the self-confirming equilibrium feature: the British type (and hence payoffs) differed from that anticipated by the Americans, although these differences come into play only once Nature removes the French threat. Nothing the British did during the generations prior to time  $t^*$  disconfirmed the Americans' beliefs about the British. Although the British believe they have the authority to intervene in the domestic affairs of the American colonies, they chose not to do so long as the French remained a threat.

Notice that the different constitutional views imply different interpretation of the British and American moves. To a 17c British type, intervention in the American colonial economy represents expropriation and opportunistic behavior. Doing so would thus alter constitution precedent throughout the empire, including in Britain itself, an outcome they would want to avoid. To a 19c British type, however, intervention represents a discretionary policy choice without constitutional implications elsewhere in the empire or Britain. In this case, intervention has no negative normative content for the British.

Nature's choice of a British type also affects the British interpretation of the American behavior. Because a 19c British type views intervention as a matter of policy, the American resistance to its intervention is treason. To the Americans, the British violation of their interpretation of the constitution is tyrannous.

The self-confirming equilibrium model thus helps explain how the two sides could cooperate for several generations, only to suddenly explode into mutual antagonism, resistance and punishment, and then revolutionary war. Moreover, this view helps explain the connection between the realm of ideas emphasized by the historians and the realm of action. The American theories of the empire's constitution reflect their attempts to understand how the British credibly committed to protect American rights in the previous era. According to this theory, British intervention – however modest – was precedent setting and thus undid the institutions protecting American liberty and rights. The British articulation of their new constitutional view exacerbated the American fears. The notion that Parliament was unlimited in its powers over the empire implied that it knew no institutional bounds on its possible intervention in America. Put simply, this said to the Americans that no credible commitments to American rights existed or were possible.

The self-confirming equilibrium view also helps explain a critical aspect of the controversy, namely that both sides believed themselves completely in the right and the other at fault. The Americans, given their constitutional view and understanding about credible commitments, saw the British intervention and new constitutional view as threatening their ways of life. They therefore sought to defend themselves. The British – facing the problems of governing a new, huge, and complex empire – sought to alter a few rules governing the Americans. They were surprised at the vehemence of the American reaction and in short order came to interpret American resistance as a move toward independence – that is, defection from the empire. They reacted harshly, helping to convince moderate Americans in the radicals' views that revolutionary action was the best path (Rakove et al., 1999).

## 5. Conclusion

In this article, we utilize the concept of self-confirming equilibrium to model a particular type of cognitive limit on rational action – the notion that individuals' mental models of their strategic environment often remain fundamentally incomplete and inaccurate. The approach requires that the players' misunderstandings cannot be arbitrary, but must instead withstand the test of time, that is, the players' misunderstandings must be compatible with the evidence resulting from their interaction along the path of the game.

The self-confirming equilibrium approach adds to our understanding of the American Revolution in several ways. First, the perspective helps clarify the history of the American Revolution, demonstrates the critical link between the realm of ideas, as emphasized by historians, and the realm of action, as occurred in the revolution. The constitutional ideals represented different understandings of the credible commitments protecting American rights and liberty. In the American view, the British behavior was tyrannous and pointed toward exploitation. As Rakove et al. (1999) emphasize, British attempts to punish American resistance – particularly the British suspension of American colonial assemblies – provided stark evidence of this perspective.

Second, and perhaps most importantly, our model explains how the two sides could cooperate peacefully for over 100 years, only then explode in war and revolution. The two sides had long held mutually inconsistent idea and ideals; yet these differences did not matter for the century prior to the Seven Years' War. Although the British believed they had the authority to alter the governance structures of the empire, they had no

interest in doing so.<sup>13</sup> Only after the war did they seek to restructure the empire, with dire consequences.

Third, beyond the differences in ideals, our perspective emphasizes the importance of the lack of common knowledge across the Atlantic: the differences in ideals emerged over a 100-year period in a way that neither perceived the other's perspective. The self-confirming component of our analysis is critical. The self-confirming equilibrium framework thus helps explain an aspect of the revolution ignored by historians: how the two sides had interacted for several generations and yet not understood one another.

The self-confirming equilibrium perspective puts the onus in part on the players' inaccurate mental models, their inability to have the complete and accurate view of their strategic situation. An outside observer, who could see both sides, as we do now, can understand how these differences emerged, were sustained, and then came into conflict. But neither side had this 'bird's eye view' perspective (Hutchins, 1995), instead interpreting the other's actions and ideas as defecting from cooperation.

#### REFERENCES

- Bailyn, Bernard (1967) *Ideological Origins of the American Revolution*. Cambridge, MA: Harvard University Press.
- Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal and Barry R. Weingast (1998) *Analytic Narratives*. Princeton, NJ: Princeton University Press.
- Blackstone, William (1765) *Commentaries on the Laws of England: A Facsimile of the First Edition of 1765–1769. Vol. 1: Of the Rights of Persons*. Chicago: University of Chicago Press.
- Camerer, Colin (1995) 'Individual Decisionmaking', in J. H. Kagel and A. Roth (eds) *Handbook of Experimental Economics*, pp. 587–703. Princeton, NJ: Princeton University Press.
- Christie, I. R. (1966) *Crisis of Empire: Great Britain and the American Colonies, 1754–1783*. New York: W. W. Norton.
- Clark, Andy (1997a) 'Economic Reason: The Interplay of Individual Learning and External Structure', in John N. Drobak and John V. C. Nye (eds) *The Frontiers of the New Institutional Economics*, pp. 269–290. San Diego, CA: Academic Press.
- Clark, Andy (1997b) *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: M.I.T. Press.
- Cook, Karen and Margaret Levi (1990) *The Limits of Rationality*. Chicago: University of Chicago Press.
- de Figueiredo, Rui J. P and Barry R. Weingast (2005) 'Self-enforcing Federalism', *Journal of Law, Economics and Organizations* 21: 103–35.
- Denzau, Arthur T. and Douglass C. North (1994) 'Shared Mental Models: Ideologies and Institutions', *Kyklos* 47: 3–31.

---

13. Recent scholarship emphasizes that the changes in the structure of the empire after the Seven Years' War can be traced earlier to events beginning before the war (see, for example, Tucker and Hendrickson, 1982). But these events do not present anything like the profound challenge to the American ideas as the explicit British intervention following the war.



- Elster, Jon (1998) 'Emotions and Economic Theory', *Journal of Economic Literature* XXXVI (March): 47–74.
- Elster, Jon (1999) *Alchemies of the Mind*. New York: Cambridge University Press.
- Frank, Robert (1985) *Choosing the Right Pond*. New York: Oxford University Press.
- Frank, Robert (1989) *Passions within Reason*. New York: W. W. Norton.
- Frey, Bruno (1997) *Not Just for the Money: An Economic Theory of Personal Motivation*. Cheltenham: Edward Elgar.
- Fudenberg, Drew and David M. Kreps (1995) 'Learning in Extensive-form Games I: Self-confirming Equilibrium', *Games and Economic Behavior* 8: 20–55.
- Fudenberg, Drew and David K. Levine (1993) 'Self-confirming Equilibrium', *Econometrica* 61: 523–45.
- Fudenberg, Drew and David K. Levine (1998) *The Theory of Learning in Games*. Cambridge, MA: M.I.T. Press.
- Greenberg, Joseph (1994) 'Social Situations without Commonality of Beliefs: Worlds Apart but Acting Together', Working Paper, McGill University.
- Greenberg, Joseph, Sudheer Gupta and Xio Luo (2004) 'Shared Actions with Divergent Perceptions: A General Game Theoretic Framework', Working Paper, McGill University.
- Greene, Jack (1986) *Peripheries and Center: Constitutional Development in the Extended Politics of the British Empire and the United States, 1607–1788*. New York: W. W. Norton.
- Greene, Jack (1991) *Negotiated Authorities: Essays in Colonial Political and Constitutional History*. Charlottesville, VA: University of Virginia Press.
- Greene, Jack (1995a) *Interpreting Early America: Historiographical Essays*. Charlottesville, VA: University of Virginia Press.
- Greene, Jack (1995b) *Understanding the American Revolution: Issues and Actors*. Charlotte: University of Virginia Press.
- Hayek, Frederich (1960) *The Constitution of Liberty*. Chicago: University of Chicago Press.
- Hutchins, Edwin (1995) *Cognition in the Wild*. Cambridge, MA: M.I.T. Press.
- Kalai, Ehud and Ehud Lehrer (1993) 'Subjective Equilibrium in Repeated Games', *Econometrica* (September) 61: 1231–40.
- Kahneman, Daniel and Amos Tversky (1979) 'Prospect Theory', *Econometrica*: 263–92.
- Lupia, Arthur and Mathew D. McCubbins (1998) *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* New York: Cambridge University Press.
- Lupia, Arthur and Natasha Zharinova (2004) 'Do Political Actors Have Beautiful Minds? Counterfactual Variation and Self-confirming Equilibria in Game Theoretic Political Science', Working Paper, University of Michigan.
- McCusker, John J. and Russell R. Menard (1985) *The Economy of British America, 1607–1789*. Chapel Hill, NC: University of North Carolina Press.
- Mansbridge, Jane J. (1990) *Beyond Self-interest*. Chicago: University of Chicago Press.
- March, James G. and Herbert A. Simon (1958) *Organizations*. New York: McGraw-Hill.
- Morgan, Edmund S. (1992 [1956]) *The Birth of the Republic, 1763–89*. Chicago: University of Chicago Press.
- North, Douglass C. (1999) 'The Process of Economic Change', Mimeo, Washington University.
- North, Douglass (2004) *Understanding the Process of Economic Change*. Princeton: Princeton University Press.
- North, Douglas C., William Summerhill and Barry R. Weingast (1999) 'Order, Disorder and Economic Change: Latin America vs. North America', in Bruce Bueno de Mesquita and Hilton Root (eds) *Governing for Prosperity*, pp. 17–57. Yale, CT: Yale University Press.
- Potter, Janice (1983) *The Liberty We Seek: Loyalist Ideology in Colonial New York and Massachusetts*. Cambridge, MA: Harvard University Press.
- Rabin, Matthew (1998) 'Psychology and Economics', *Journal of Economic Literature* 36: 11–46.



- Rakove, Jack N. (1979) *The Beginnings of National Politics: An Interpretive History of the Continental Congress*. Baltimore, MD: Johns Hopkins University Press.
- Rakove, Jack N. (1996) *Original Meanings: Politics and Ideas in the Making of the Constitution*. New York: Knopf.
- Rakove, Jack N., Andrew Rutten and Barry R. Weingast (1999) 'Ideas, Interests, and Credible Commitments in the American Revolution', Working Paper, Hoover Institution.
- Reid, John P. (1995) *Constitutional History of the American Revolution*. Madison, WI: University of Wisconsin Press.
- Sargent, Thomas J. (1999) *The Conquest of American Inflation*. Princeton, NJ: Princeton University Press.
- Satz, D. and J. Ferejohn (1994) 'Rational Choice and Social Theory', *Journal of Philosophy* 91(2): 71–87.
- Simon, Herbert A. (1947) *Administrative Behavior*. New York: MacMillan.
- Simon, Herbert (1947) *Administrative Behaviour*. New York: MacMillan.
- Smith, Joseph H. (1965) *Appeals to the Privy Council from the American Plantations*. New York: Octagon Books.
- Tucker, Robert W. and David C. Hendrickson (1982) *The Fall of the First British Empire: Origins of the War of American Independence*. Baltimore, MD: Johns Hopkins University Press.
- Weingast, Barry R. (1995) 'The Economic Role of Political Institutions: Market-Preserving Federalism and Economic Development', *Journal of Law, Economics, and Organization*: 1–31.
- Wood, Gordon S. (1966) 'Rhetoric and Reality in the American Revolution', *William and Mary Quarterly* 23: 3–32.
- Wood, Gordon S. (1969) *The Creation of the American Republic, 1776–1787*. New York: W. W. Norton.

---

RUI J. P. DE FIGUEIREDO JR is Associate Professor at the Haas School of Business and Department of Political Science at University of California, Berkeley. He is a co-author of *The Outsider: Prejudice and Politics in Italy* and author or co-author of several articles on American politics and political economy in *American Political Science Review*, *American Journal of Political Science*, *Journal of Law and Economics*, and other journals.

JACK RAKOVE is William Robertson Coe Professor of History and American Studies at Stanford University and Senior Fellow at the Hoover Institution. He is the author of *Original Meanings: Politics and Ideas in the Making of the Constitution* and other books and articles on American history and politics.

BARRY R. WEINGAST is the Ward C. Krebs Family Professor in the Department of Political Science at Stanford University and a senior fellow at the Hoover Institution. He is the author and co-author of many articles on American politics and political economy in leading political science and economics journals. He is a co-editor (with Donald Wittman) of the *Oxford Handbook of Political Economy* (2006). ADDRESS: Department of Political Science, Encina Hall West, Room 100, Stanford University, Stanford, CA 94305-6044, USA [email: weingast@stanford.edu].