2001-16

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

LYING FOR STRATEGIC ADVANTAGE: RATIONAL AND BOUNDEDLY RATIONAL MISREPRESENTATION OF INTENTIONS

BY

VINCENT P. CRAWFORD

DISCUSSION PAPER 2001-16
SEPTEMBER 2001

LYING FOR STRATEGIC ADVANTAGE: RATIONAL AND BOUNDEDLY RATIONAL MISREPRESENTATION OF INTENTIONS

by Vincent P. Crawford*

24 September 2001

"Lord, what fools these mortals be!"—Puck, *A Midsummer Night's Dream*, Act 3

"You may fool all the people some of the time; you can even fool some of the people all the time; but you can't fool all of the people all the time."—Abraham Lincoln

"Now give Barnum his due."—John Conlisk (2001)

Abstract

Starting from Hendricks and McAfee's (2000) example of the Allies' decision to feint at Calais and attack at Normandy on D-Day, this paper models misrepresentation of intentions to competitors or enemies. Allowing for the possibility of bounded strategic rationality and rational players' responses to it yields a sensible account of lying via costless, noiseless messages. In many cases the model has generically unique pure-strategy sequential equilibria, in which rational players exploit boundedly rational players, but are not themselves fooled. In others, the model has generically essentially unique mixed-strategy sequential equilibria, in which rational players' strategies protect all players from exploitation.

Keywords: lying, misrepresentation of intentions, preplay communication, noncooperative games, conflict (*JEL* C72, D72, D80)

*Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, California, 92093-0508. I am grateful to John Conlisk, Miguel Costa-Gomes, Herb Newhouse, Joel Sobel, and Kang-Oh Yi for helpful advice; and to the National Science Foundation for research support.

Lying for strategic advantage about planned actions, or *intentions*, is a common feature of economic and political as well as military life. Such lying frequently takes the extreme form of active misrepresentation, as opposed to less than full, honest disclosure. Examples range from the University of California's three (!) consecutive "last chance" voluntary early retirement incentive programs in the early 1990s; ex-President George Bush's regrettably memorable 1988 campaign promise, "Read my lips: no new taxes"; the universal practice of lying about planned currency devaluations; Nathan Rothschild's pretense of having received early news of a British defeat at Waterloo; and Hitler's 1939 non-aggression pact with Stalin.[1] In other cases, the effects of active misrepresentation are duplicated by tacitly exploiting widespread misperceptions, as in accelerationist monetary policy; periodic but unpredictable temporary investment tax credits, or regularizations of the status of illegal immigrants; the failure to disclose known product safety hazards; and deceptive accounting practices in the private and the public sector.[2]

These examples share two common features. All involve misrepresentation via agreements, statements, or non-statements that in themselves have little or no direct costs. And all involve situations in which the parties have predominantly conflicting interests, so that successful deception benefits the deceiver only at the expense of the deceived. Nonetheless, the misrepresentation often succeeds. In fact, in many of the examples the public has so completely internalized the logic of misrepresentation that criticism of the gullibility of those deceived is as common as criticism of the misrepresentation.[3]

Theory lags behind the public's intuition. The examples' common features suggest that, to a first approximation, they can be modeled as communication via costless messages ("cheap talk") in a zero-sum two-person game. But in such a model costless messages must be ignored in equilibrium: If a player could benefit by responding to the

[1]Roland Benabou and Guy Laroque (1992) give several interesting examples concerning lying to manipulate financial markets, including Rothschild's, which allowed him to make large clandestine purchases of British government securities at depressed prices. Examples of lying in international politics are easy to find, and it is probably no accident that there is a board game called Diplomacy in which success depends on forming unenforceable agreements with other players and being the first to break them.
[2]Paul Krugman (2001) discusses the current Bush administration's use of "creative" accounting to make the 2001 tax cut appear feasible without dipping into the Social Security surplus, an example that was highly topical before the question was made moot by the terrorist attacks of 11 September 2001.
[3]Mike Royko's (1988) prescient view of Bush's "Read my lips" promise is an entertaining example. An official who does not lie about his country's plan to devalue its currency risks being removed from office, if not institutionalized. Other examples are explicitly covered by proverbs such as "All's fair in love and war."

other player's message, his response would hurt the other player, who would therefore do better to make his message uninformative. Thus, in equilibrium no information is conveyed by the message, but neither is anyone fooled by it.[4]

This result is appealing in its simplicity, but it leaves us with no systematic way to think about an important phenomenon. This paper seeks to provide one by proposing a way to model bounded strategic rationality and rational players' responses to it, and using it to analyze a simple model of misrepresentation of intentions to competitors or enemies.

My analysis is inspired in part by Kenneth Hendricks and R. Preston McAfee's (2000; henceforth "HM") analysis of misrepresentation of intentions via what they call "feints" or "inverted signaling." HM's primary interest, like mine, is in economic and political applications; but they motivate their analysis by Operation Fortitude, the Allies' successful attempt to mislead the Germans about their intention to land in Normandy rather than the Pas de Calais (the obvious choice ex ante) on D-Day, 6 June 1944, and I will follow them in this.[5] Their model is a zero-sum two-person game. First the attacker chooses (possibly randomly) between two possible locations and allocates a fixed budget of force between them. Next, the defender privately observes a binary signal whose probability distribution depends on the attacker's allocation, and allocates (possibly randomly) his own budget of force between the two locations. The attack location and players' force allocations then determine their payoffs. The attacker's allocation is like a noisy message to the defender, but as in other models of costly signaling, its large direct payoff implications sometimes allow equilibria in which it is not ignored.

HM assume that the payoff function and the conditional probability distribution of the signal are both symmetric across locations. They show, under plausible additional assumptions, that equilibrium in their game must involve some attempt by the attacker to misrepresent his intentions (allocating force to both locations with positive probability)

---

[4]See Vincent Crawford and Joel Sobel's (1982) and Joseph Farrell's (1993) analyses of strategic communication of private information and intentions. Crawford and Sobel's equilibria have no active misrepresentation, only intentional vagueness, taking the extreme form of no transmission if the Sender's and Receiver's preferences differ enough to make the game effectively zero-sum; Farrell coined the term "babbling" for such equilibria. Farrell and Matthew Rabin (1996) and Crawford (1998) survey the theory.

[5]The deception was so successful that the Germans kept 19 divisions in Calais for several critical days after D-Day. HM summarize the history; see also Gordon Harrison (1951, especially Appendix A) and Anthony Kemp (1994). HM also give several examples of economic and political misrepresentation, in which firms

and that his attempt succeeds (inducing the defender to allocate force to both locations with positive probability). For, if the defender ignored his observation of the signal, the attacker would assign all of his force to his intended attack location; but if the defender anticipated this, the attacker would prefer to allocate some force to the other location.

HM identify equilibria in their model in two cases, distinguished by the signal's informativeness. When the signal is not very informative, they identify "full-defense" equilibria, in which the attacker deterministically allocates most of his force to one attack location but randomizes the location itself, and the defender allocates all of his own force deterministically, to the location the signal suggests is more likely to be attacked. When the signal is more informative, they identify "split-defense" equilibria, in which the attacker randomizes his allocation and attack location in such a way that the defender can draw no inference from the signal, and the defender also randomizes his allocation. In these equilibria, with positive probability the attacker allocates more than half his force to the location he does *not* attack. HM also obtain intriguing comparative statics results, showing that when the signal is not very informative a reduction in noise hurts the attacker; but that when it is more informative, a reduction in noise benefits the attacker.

HM stress that their explanation of misrepresentation depends on the noisiness of the signal: "With perfect observability, feints differ from the standard analysis in inconsequential ways. In particular, were the Germans to observe the actual allocation of allied forces, it would not have been possible for the Allies to fool the Germans. Thus, imperfect observation is a critical element for modeling feints."

HM's analysis makes significant progress in understanding the phenomenon of misrepresentation, but it has three troubling aspects. I shall describe them from the point of view of Operation Fortitude, although they are equally troubling in other applications.

First, the cost to the Allies of faking the preparations for an invasion of Calais was small compared to that of the preparations for the actual invasion of Normandy, hence more like cheap talk than HM's identification of feints with sizeable fractions (sometimes more than half) of the attacker's force would suggest. The Germans knew as well as the Allies that it was feasible to fake, or conceal, invasion preparations at no great cost. In a

---

distort their exploration or bidding strategies to mislead competitors, political candidates campaign in areas they believe are unimportant to divert their opponents' efforts from areas they consider crucial, and so on.

standard equilibrium analysis, they would then rationally ignore both the faked evidence that the attack would be at Calais and the lack of evidence that it would be at Normandy. But they didn't—and Allied planners didn't expect them to, with anything like certainty.

Second, HM's analysis does not reflect the asymmetry between Normandy and Calais that is arguably the most salient feature of Operation Fortitude.[6] Why not feint at Normandy and attack at Calais instead, particularly if the deception has a fair chance of success? Allied planners rejected Calais in favor of Normandy early in their planning, mainly (but not entirely) because the proximity to England that made it the obvious attack location was also obvious to the Germans, who were expected to defend it so heavily that on balance, Normandy would be preferable (Harrison (1951)). Neither Allied planners' choice of Normandy nor the fact that they did not explicitly randomize it is inconsistent with HM's equilibria per se, because they assign positive probabilities to both attack locations and in a mixed-strategy equilibrium in beliefs, a player need not bear any uncertainty about his own decision (Robert Aumann and Adam Brandenburger (1995)). But Allied planners were not indifferent between the locations, and an explanation that treats their choice as an accidental feature of the history may miss something important.

Finally, an analysis of equilibrium in a game without precedent—of which Operation Fortitude and D-Day are perhaps the quintessential example—implicitly rests on the assumption that players' rationality and beliefs are at least mutual knowledge (Aumann and Brandenburger (1995, Theorem A)). These assumptions are more than usually strained in HM's model, whose equilibria involve a delicate balance of wholly or partly mixed strategies that depend on the details of the signal distribution.

This paper shows that a sensible account of misrepresentation of intentions can be given, in a simpler game, and with costless and noiseless messages, by allowing for the possibility of bounded strategic rationality. The model and analysis fully reflect the low message costs, the importance of payoff asymmetry across actions, and the difficulty of justifying a delicate equilibrium analysis of a game without precedent just noted.

The model is based on the class of zero-sum two-person perturbed Matching Pennies games in Figure 1. Two players, a Sender (analogous to the Allies) and a

[6]HM"s only reference to the asymmetry is to note that when their signal is not very informative, if the attacker's payoffs make one location easier to attack, that location is *more* likely to be attacked.

Receiver, choose simultaneously between two pure actions, U for Up (analogous to attacking at Calais) or D for Down for the Sender and L for Left (analogous to defending Normandy) and R for Right for the Receiver. I assume throughout that $a > 1$, reflecting the lesser difficulty of an unanticipated invasion of Calais. Before playing this *underlying game*, the Sender sends the Receiver a costless, non-binding, noiseless message, u or d, about his intended action, with u (d) representing action U (D) in a commonly understood language (Farrell (1993)). Players then choose their actions simultaneously. The structure of the game is common knowledge. These games differ from HM's in having costless and noiseless messages separate from the attacker's force allocation, simultaneous, 0-1 allocations of force to locations, and a payoff asymmetry across actions.[7]

		Receiver	
		Left	**Right**
Sender	**Up**	$-a$, a	0, 0
	Down	0, 0	-1, 1

Figure 1. The underlying game

In a standard equilibrium analysis of this game, in any equilibrium (subgame-perfect or not) the Sender's message must be uninformative, in that the probability that he plays U conditional on his message is independent of the message; and the Receiver must ignore it, in the sense that the probability that he plays L is independent of the Sender's message.[8] The underlying game must therefore be played according to its unique mixed-strategy equilibrium, in which the Sender plays U with probability $1/(1+a)$ and the Receiver plays L with probability $1/(1+a)$, with respective expected payoffs $a/(1+a)$ and $-a/(1+a)$.[9] Thus, communication is ineffective and misrepresentation cannot occur.

[7]Because each side's forces were actually somewhat dispersed, the discrete force allocations in the present model should be thought of as representing *principal* attack or defense locations.

[8]The Sender can make his message uninformative by choosing a strategy in which he always sends the same message, or a strategy in which he randomizes his message independently of his action.

[9]This equilibrium illustrates a strategic principle noted by John von Neumann and Oskar Morgenstern (1953 [first edition 1944 (!)], pp. 175-176) in that, counterintuitively, the Sender's probability of playing U is, like the Receiver's probability of playing L, a decreasing function of a. Thus, in Operation Fortitude, the

The closest precedents for a non-equilibrium analysis of this kind of game are Farrell (1988) and Rabin (1994), who study preplay communication about intentions via cheap talk, mainly in games in which players have substantial common interests, using augmented notions of rationalizability.[10] The model proposed here is similar in spirit, but it relaxes the assumption of equilibrium in a different way, imposing more structure on players' behavior by allowing for the possibility that they use simple, boundedly rational decision rules. The analysis is otherwise completely standard.

Specifically, I assume that each player role is filled randomly from a separate distribution of decision rules, or *types*, that assigns positive probability to certain boundedly rational, or *Mortal*, types as well as to a *Sophisticated* type that is a natural extension of the ideal of a fully strategically rational player to this setting. Players do not observe each other's types, but the structure of the game, including the type distributions, is common knowledge. *Sophisticated* players satisfy the usual mutual knowledge of beliefs and rationality assumptions with respect to each other, and they can also use their knowledge of the structure to predict the probability distributions of *Mortal* players' strategies.[11] *Mortal* players can be thought of as rational expected-payoff maximizers, if desired, but their beliefs about others' strategies generally differ from equilibrium beliefs.

The possibility of interacting with *Mortal* players fundamentally alters the game from *Sophisticated* players' point of view. Because *Mortal* players' strategies are determined independently of each other's and *Sophisticated* players' strategies, as explained below, they can be treated as exogenous, allowing the analysis to focus on a *reduced game* between possible *Sophisticated* players in each role.[12] In this game, a *Sophisticated* Sender's incentives to misrepresent his intentions weigh the equilibrium

---

Germans would ignore all messages and be more likely to defend Calais than Normandy. Crawford and Dennis Smallwood (1984) analyze the comparative statics of such payoff changes in general two-person zero-sum games with mixed-strategy equilibria, identifying the general principle that underlies this result.

[10]See also Miguel Costa-Gomes (2002), who extends Rabin's analysis to interpret experimental data.

[11]The *Mortal* types are adapted to communication games from the types based on iterated best responses in Dale Stahl and Paul Wilson (1995) and Costa-Gomes, Crawford, and Bruno Broseta (2001). Andreas Blume et al. (1998, 2001) find evidence of some such types in experiments with communication games. The *Sophisticated* types are adapted from the *Sophisticated* type in Costa-Gomes et al. (2001).

[12]An equilibrium analysis of the reduced game generalizes standard equilibrium analysis of the original game, in that common knowledge that all players are *Sophisticated* would make their beliefs common knowledge and therefore the same, and so in equilibrium (Aumann and Brandenburger (1995)). Similar techniques are used by Colin Camerer, Teck-Hua Ho, and Juin-Kuan Chong (2002) to analyze a reduced game with sophisticated and adaptive learners, elucidating a phenomenon they call "strategic teaching."

response of a *Sophisticated* Receiver against those of various types of *Mortal* Receivers, of whom some invert the Sender's message and others believe it. Thus the model adopts a view of human nature close to Lincoln's, which is more nuanced and arguably more realistic than Puck's. The reduced game is no longer zero-sum, because a *Sophisticated* player's expected payoffs are influenced by possible interaction with a *Mortal* opponent, whose payoff may differ from a *Sophisticated* opponent's. Its messages are no longer cheap talk, because a *Sophisticated* Sender's message directly influences his expected payoff via *Mortal* Receivers' responses. Finally, the reduced game has incomplete information, and a *Sophisticated* Receiver can sometimes draw inferences from a Sender's message about his type.[13] These differences suggest that an analysis with some *Mortal* players may differ in interesting ways from a standard equilibrium analysis.

As one might expect, most features of the model's equilibria depend on the relative frequencies of *Sophisticated* and *Mortal* players in each role, but the possibility of *Mortal* players affects *Sophisticated* players' behavior in perhaps unexpected ways.

When the probabilities of a *Sophisticated* Sender and Receiver are high relative to the payoff advantage of an unanticipated attack at Calais over one at Normandy, the reduced game has a generically essentially unique sequential equilibrium in mixed strategies, similar to the standard analysis's babbling message followed by mixed-strategy equilibrium in the underlying game. In this case *Sophisticated* players' equilibrium mixed strategies offset each other's gains from fooling *Mortal* players, and in each role, *Sophisticated* players have the same expected payoffs as their *Mortal* counterparts. Further, all types' expected payoffs are the same as in the standard analysis.

By contrast, when the probabilities of a *Sophisticated* Sender and Receiver are low relative to the payoff advantage of Calais, the reduced game has a generically unique sequential equilibrium in pure strategies.[14] In these equilibria a *Sophisticated* Receiver

[13]Only *Sophisticated* players play an active strategic role, but *Mortal* players' types, which determine their strategies, influence *Sophisticated* players' equilibrium strategies and welfare. The role of *Mortal* players in the analysis resembles Philip Reny's (1992) notion of explicable equilibrium, which models "trembles" via "complete theories"; here, however, the trembles are not eliminated by passing to the limit. The preceding observations suggest that the analysis is robust to small-to-moderate violations of its basic assumptions.

[14]Generic uniqueness of sequential equilibrium is unusual in signaling games, which normally have both essential nonuniqueness and, in the case of cheap-talk games, inessential nonuniqueness due to the ambiguity of the meaning of costless messages in equilibrium (Crawford and Sobel (1982)). My analysis avoids this ambiguity because I assume that *Mortal* Receivers understand the literal meanings of the

can predict a *Sophisticated* Sender's action, and vice versa; thus their communication is "disciplined" in the sense of Farrell and Robert Gibbons (1989). In the present model, however, the discipline comes from the implicit presence of *Sophisticated* players' *Mortal* alter egos, rather than from real other players. Further, the Sender's message plays a different role, conveying information about the Sender's type rather than his intentions.

When the probability of a *Sophisticated* Sender is relatively low and the probability of a *Mortal* Receiver who believes the Sender's message is not too high, the model has a unique sequential equilibrium in which a *Sophisticated* Sender sends message u but plays D—like feinting at Calais and attacking at Normandy—and both a *Sophisticated* Receiver and a *Mortal* Receiver who believes the Sender's message play R—like defending Calais. In such an equilibrium, a *Sophisticated* Receiver plays R because in the parameter configurations that support it, being "fooled" at a unit cost of 1 by a *Sophisticated* Sender is preferable to being "fooled" at a unit cost of *a* by both kinds of *Mortal* Sender. There are also configurations with unique sequential equilibria in which a *Sophisticated* Sender sends message d but plays U, but it is argued below that the conditions for the equilibria that resemble Operation Fortitude are more realistic.

The explanation of Operation Fortitude these equilibria suggest is less subtle, and perhaps more credible, than HM's explanation: *Sophisticated* Allied planners (or *Mortal* planners who make a point of lying to Germans) conceal their preparations for invading Normandy and fake preparations for invading Calais, knowing that the cost of faking is low; that the Germans may be the type of *Mortal* who can be fooled this way; and that even *Sophisticated* Germans prefer to defend Calais because they think the Allies are probably *Mortal*, and if so will attack at Calais. *Mortal* Germans who believe the Allies' messages are fooled because they are too literal-minded (or perhaps too clever) to see through the deception.[15] *Sophisticated* Germans see through the deception, but still prefer to defend Calais, even at the risk of being "fooled" by possibly *Sophisticated* Allies.

Sender's messages and other players know this. It avoids essential nonuniqueness because *Mortal* Senders ensure that both messages have positive probability, and Senders' and Receivers' interests are opposed.

[15]As explained below, *Mortal* Germans are fooled if they believe they are an even number of steps ahead of the Sender, in the hierarchy of iterated-best-response types. Other types of *Mortal* German are not fooled because they believe they are an odd number of steps ahead; but *Sophisticated* Allied planners know that (in this case) such types are less likely than *Mortal* Germans who will be fooled by feinting at Normandy.

Importantly for applications, in this explanation *Sophisticated* players' sequential equilibrium strategies (pure or mixed) depend only on the payoffs and population parameters that reflect simple, portable facts about human behavior that could be learned in many conflict situations, without regard to the quality of the analogy with the present situation (fortunately, for applications as unprecedented as Operation Fortitude). Further, in all the model's pure-strategy sequential equilibria, *Sophisticated* players' strategies are their unique extensive-form rationalizable strategies, identifiable by at most three steps of iterated conditional dominance (Makoto Shimoji and Joel Watson (1998)).

With regard to welfare, *Sophisticated* players in either role by definition do at least as well in equilibrium as their *Mortal* counterparts. In the mixed-strategy sequential equilibria that arise when the probabilities of a *Sophisticated* Sender and Receiver are both relatively high, in each role *Sophisticated* and *Mortal* players have the same expected payoffs: Perhaps surprisingly, the prevalence of *Sophisticated* players fully protects *Mortal* players from exploitation. By contrast, in pure-strategy sequential equilibria (and in some mixed-strategy sequential equilibria), *Sophisticated* players in either role do strictly better than their *Mortal* counterparts. Their advantage comes from their ability to avoid being fooled (except by choice, when it is the lesser of two evils) and from *Sophisticated* Senders' ability to choose which type(s) of Receiver to fool.

These results suggest that an adaptive analysis of the dynamics of the type distribution, in the style of Conlisk (2001), would show that *Sophisticated* and *Mortal* players can coexist in long-run equilibrium whether or not *Sophisticated* players have higher costs, justifying the assumptions about the type probabilities maintained here.[16]

The rest of the paper is organized as follows. Section I completes the specification of the model by describing the behavior of *Mortal* types and showing how to construct the reduced game. Section II characterizes the model's sequential equilibria, showing how they depend on the payoffs and the type distribution. Section III compares *Mortal* and *Sophisticated* Sender and Receiver types' equilibrium welfares and briefly discusses an adaptive model of the evolution of the type distribution. Section IV discusses related work, and Section V is the conclusion.

[16]I am grateful to Kang-Oh Yi for this observation.

This section completes the specification of the model by describing *Mortal* types' behavior, and shows how to construct the reduced game between *Sophisticated* players.

A Sender's pure strategies are (message, action|sent u, action|sent d) = (u,U,U), (u,U,D), (u,D,U), (u,D,D), (d,U,U), (d,U,D), (d,D,U), or (d,D,D); and a Receiver's are (action|received u, action|received d) = (L,L), (L,R), (R,L), or (R,R). Table 1 lists some plausible *Mortal* Sender and Receiver types, with their *Sophisticated* counterparts.[17]

Sender type	**Behavior (b.r. ≡ best response)**	**message, action|sent u, action|sent d**
Credible ≡ *W0*	tells the truth	u,U,D
W1 (*Wily*)	lies (b.r. to *S0*)	d,D,U
W2	tells truth (b.r. to *S1*)	u,U,D
W3	lies (b.r. to *S2*)	d,D,U
Sophisticated	b.r. to population	depends on the type probabilities
Receiver type	**Behavior**	**action|received u, action|received d**
Credulous ≡ *S0*	believes (b.r. to *W0*)	R, L
S1 (*Skeptical*)	inverts (b.r. to *W1*)	L, R
S2	believes (b.r. to *W2*)	R, L
S3	inverts (b.r. to *W3*)	L, R
Sophisticated	b.r. to population	depends on the type probabilities

Table 1. *Mortal* and *Sophisticated* Sender and Receiver types

Like most boundedly rational strategic decision rules, these *Mortal* types use step-by-step procedures that generically determine unique, pure strategies, and avoid simultaneous determination of the kind used to define equilibrium. In the words of Reinhard Selten (1998, p. 433), "Basic concepts in game theory are often circular in the sense that they are based on definitions by implicit properties….Boundedly, [sic] rational strategic reasoning seems to avoid circular concepts. It directly results in a procedure by which a problem solution is found. Each step of the procedure is simple, even if many case distinctions by simple criteria may have to be made." *Mortal* players' strategies are therefore determined independently of each other's and *Sophisticated* players' strategies.

This independence allows a simple characterization of the implications of bounded rationality. Restricting attention to the *Mortal* types in Table 1 for definiteness,

note that a *Wily* Sender, *Wj*, with *j* odd always lies; I lump these *Mortal* Sender types together under the heading *Liars* from now on. A Wily Sender with *j* even (including *Credible* as an honorary *Wily* type, *W0*) always tells the truth; I lump these types together under the heading *Truthtellers*. Similarly, a *Skeptical* Receiver, *Sk*, with *k* odd always inverts the Sender's message, and a *Skeptical* Receiver with *k* even (including *Credulous* as an honorary *Skeptical* type, *S0*) always believes it; I lump these *Mortal* Receiver types together under the headings *Inverters* and *Believers*.[18] Thus, the behavior of a Sender population can be summarized by $s_l \equiv$ Pr{Sender is a *Liar*}, $s_t \equiv$ Pr{Sender is a *Truthteller*}, and $s_s \equiv$ Pr{Sender is *Sophisticated*}, where $s_l + s_t + s_s = 1$; and the behavior of a Receiver population can be summarized by $r_i \equiv$ Pr{Receiver is an *Inverter*}, $r_b \equiv$ Pr{Receiver is a *Believer*}, and $r_s \equiv$ Pr{Receiver is *Sophisticated*}, where $r_i + r_b + r_s = 1$.

To avoid trivialities, I assume that these type probabilities are all strictly positive in both populations. I also ignore nongeneric parameter configurations, and all "if and only if" (henceforth "iff") statements should be interpreted in the generic sense.

Table 1 shows that, although *Inverters* and *Believers* always choose different actions for a given message, its *Mortal* Sender types always play U on the equilibrium path. This conclusion follows naturally from the fact that U yields the Sender higher expected payoffs, other things equal; but it does not hold for all conceivable boundedly rational Sender types. Nonetheless, I shall maintain it as a plausible simplifying assumption; moderate violations would not significantly alter the results.

Because all *Mortal* Senders play U on the equilibrium path, *Liars* always send message d and *Truthtellers* send message u. Thus, both messages always have positive probability, and a *Sophisticated* Sender is always pooled with one *Mortal* Sender type.

After receiving a message for which a *Sophisticated* Sender's strategy specifies playing U with probability 1, like *Mortal* Senders, a *Sophisticated* Receiver's best response is R. But otherwise his best response may depend on his posterior probability or *belief*, *z*, that the Sender is *Sophisticated*. If *x* is the message and *y* is a *Sophisticated*

---

[17]I assume for convenience that *Credible* Senders play u,U,D rather than d,U,D, even though both strategies are truthful and both yield the Sender the same payoff, 0, if his message is always believed. *Credible* Senders could be given a strict preference for u,U,D by perturbing payoffs slightly.

[18]That such types can be lumped together in this way illustrates a kind of paradox of bounded strategic rationality, in that with a finite number of possibilities for guessing and outguessing, it is as bad to be too

Sender's probability of sending message u, a *Sophisticated* Receiver's belief is determined by Bayes' Rule: $z \equiv f(x,y)$, where $f(u,y) \equiv ys_s/(s_t+ys_s)$ and $f(d,y) \equiv (1–y)s_s/[(1–y)s_s+s_l]$.

		Receiver			
		L,L	L,R	R,L	R,R
Sender	u,U,U	$a(r_i+r_s)$, $–a$	$a(r_i+r_s)$, $–a$	ar_i, 0 **A**	ar_i, 0 **B**
	u,U,D	$a(r_i+r_s)$, $–a$	$a(r_i+r_s)$, $–a$	ar_i, 0 **A'**	ar_i, 0 **B'**
	u,D,U	r_b, $–as_t/(s_s+s_t)$	r_b, $–as_t/(s_s+s_t)$	(r_b+r_s), $–s_s/(s_s+s_t)$	(r_b+r_s), $–s_s/(s_s+s_t)$ **Γ**
	u,D,D	r_b, $–as_t/(s_s+s_t)$	r_b, $–as_t/(s_s+s_t)$	(r_b+r_s), $–s_s/(s_s+s_t)$	(r_b+r_s), $–s_s/(s_s+s_t)$ **Γ '**
	d,U,U	$a(r_b+r_s)$, $–a$	ar_b, 0 **Δ**	$a(r_b+r_s)$, $–a$	ar_b, 0 **E**
	d,U,D	r_i, $–as_l/(s_s+s_l)$	(r_i+r_s), $–s_s/(s_s+s_l)$	r_i, $–as_l/(s_s+s_l)$	(r_i+r_s), $–s_s/(s_s+s_l)$ **Z**
	d,D,U	$a(r_b+r_s)$, $–a$	ar_b, 0 **Δ '**	$a(r_b+r_s)$, $–a$	ar_b, 0 **E'**
	d,D,D	r_i, $–as_l/(s_s+s_l)$	(r_i+r_s), $–s_s/(s_s+s_l)$	r_i, $–as_l/(s_s+s_l)$	(r_i+r_s), $–s_s/(s_s+s_l)$ **Z'**

Figure 2. Payoff matrix of the reduced game between a *Sophisticated* Sender and Receiver

		Receiver	
		L	**R**
Sender	**U**	$–a$ $a(r_i+r_s)$	0 ar_i
	D	$–a(1–z)$ r_b	$–z$ (r_b+r_s)

Figure 3a. "u" game following message u

		Receiver	
		L	**R**
Sender	**U**	$–a$ $a(r_b+r_s)$	0 ar_b
	D	$–a(1–z)$ r_i	$–z$ (r_i+r_s)

Figure 3b. "d" game following message d

Figure 2 gives the payoff matrix of the reduced game between a *Sophisticated* Sender and Receiver, using these observations to derive *Sophisticated* players' expected payoffs. If, for example, a *Sophisticated* Sender's strategy is u,U,D and a *Sophisticated* Receiver's strategy is R,L, the former plays U and the latter plays R when he receives message u. Thus, all Sender types play U, *Inverters* play L, *Believers* and *Sophisticated* Receivers play R, a *Sophisticated* Sender's expected payoff is ar_i, and a *Sophisticated* Receiver's is 0. If, instead, a *Sophisticated* Sender's strategy is u,D,U and a *Sophisticated*

---

much wilier, or more skeptical, than one's opponent as to be too much less wily, or skeptical. By contrast,

Receiver's strategy is L,R, the former plays D and the latter plays L when he receives message u. All other Sender types play U, *Inverters* play L, and *Believers* play R. A *Sophisticated* Sender's expected payoff is r_b; and a *Sophisticated* Receiver's, whose posterior belief that the Sender is *Mortal* is $1- s_s/(s_t + s_s) \equiv s_t/(s_t + s_s)$, is $-as_t/(s_t + s_s)$.

Figure 3 gives the payoff matrices of the reduced "u" and "d" games following messages u and d, as determined by a *Sophisticated* Receiver's belief, $z \equiv f(x,y)$.

Because messages are costless, the only difference between type populations in which the frequencies of *Mortal* Senders and Receivers are interchanged is which message fools which type. Figure 2 reflects this symmetry, in that simultaneous permutations of the probabilities of *Liars* and *Truthtellers*, and of *Believers* and *Inverters*, yield an equivalent game. Figure 3's u and d games are identical except for interchanged roles of r_i and r_b, because they differ only in whether *Inverters* or *Believers* are fooled.

II. Analysis

In this section I characterize the sequential equilibria of the reduced game, as functions of the payoff a and the type probabilities. Sequential equilibrium combines the standard notion of sequential rationality with consistency restrictions on players' beliefs.[19]

(E) d,U,U; R,R iff $r_b > r_i$, $ar_b + r_i > 1$, and $r_i > 1/(1+a)$ (true iff $r_b > r_i > 1/(1+a)$)
(E') d,D,U; R,R iff $r_b > r_i$, $ar_b + r_i > 1$, and $r_i < 1/(1+a)$
(Γ) u,D,U; R,R iff $r_b > r_i$, $ar_b + r_i < 1$, $r_b > 1/(1+a)$, and $s_s < as_t$
(Γ_m) m,D,U; R,R iff $r_b > r_i$, $ar_b + r_i < 1$, $r_b > 1/(1+a)$, and $s_s > as_t$
(Γ') u,D,D; R,R iff $r_b > r_i$, $ar_b + r_i < 1$, $r_b < 1/(1+a)$, and $s_s < as_t$ (true iff $r_i < r_b < 1/(1+a)$)
(Γ'_m) m,M_u,M_d; M_u,M_d iff $r_b > r_i$, $ar_b + r_i < 1$, $r_b < 1/(1+a)$, and $s_s > as_t$
(B) u,U,U; R,R iff $r_i > r_b$, $ar_i + r_b > 1$, and $r_b > 1/(1+a)$ (true iff $r_i > r_b > 1/(1+a)$)
(B') u,U,D; R,R iff $r_i > r_b$, $ar_i + r_b > 1$, and $r_b < 1/(1+a)$
(Z) d,U,D; R,R iff $r_i > r_b$, $ar_i + r_b < 1$, $r_i > 1/(1+a)$, and $s_s < as_l$
(Z_m) m,U,D; R,R iff $r_i > r_b$, $ar_i + r_b < 1$, $r_i > 1/(1+a)$, and $s_s > as_l$
(Z') d,D,D; R,R iff $r_i > r_b$, $ar_i + r_b < 1$, $r_i < 1/(1+a)$, and $s_s < as_l$ (true iff $r_b < r_i < 1/(1+a)$)
(Z'_m) m,M_u,M_d; M_u,M_d iff $r_i > r_b$, $ar_i + r_b < 1$, $r_i < 1/(1+a)$, and $s_s > as_l$

Table 2. Sequential equilibria of the reduced game between a *Sophisticated* Sender and Receiver[20]

in Conlisk's (2001) model *Tricksters* always find a way to outwit *Suckers*, just as Puck does with mortals.

[19]Under my assumptions both messages always have positive probability, so zero-probability updating is not an issue, and any notion that captures the idea of sequential rationality would yield the same results.

The Greek capital letters in Figure 2 identify the strategy combinations for *Sophisticated* players that are pure-strategy equilibria of the reduced game (sequential or not) for some parameter configurations. Table 2 lists the sequential equilibria for those configurations, and Figure 4 [end of the paper] graphs the configurations in (r_i,r_b)-space.

Proposition 1, proved in the Appendix, is the basic characterization result:

PROPOSITION 1: Unless either $r_b > r_i$, $ar_b + r_i < 1$, and $s_s > as_t$, or $r_i > r_b$, $ar_i + r_b < 1$, and $s_s > as_l$, the reduced game has a generically unique sequential equilibrium in pure strategies, in which a *Sophisticated* Sender's and Receiver's strategies are as given in Table 2 and Figure 4. In these sequential equilibria, a *Sophisticated* Receiver's strategy is R,R; and a *Sophisticated* Sender plays U (D) on the equilibrium path iff $a \max\{r_b,r_i\} + \min\{r_b,r_i\} > (<) 1$ and sends message d (u) iff $r_b > (<) r_i$. *Sophisticated* players' sequential equilibrium strategies are their unique extensive-form rationalizable strategies, identifiable by at most three steps of iterated conditional dominance.

If, instead, either (i) $r_b > r_i$, $ar_b + r_i < 1$, and $s_s > as_t$; or (ii) $r_i > r_b$, $ar_i + r_b < 1$, and $s_s > as_l$, the reduced game has a generically unique or essentially unique mixed-strategy sequential equilibrium, in which a *Sophisticated* Sender's and Receiver's strategies are as given in Table 2 and Figure 4. In case (i), if $r_b < 1/(1+a)$, there are multiple mixed-strategy sequential equilibria, in each of which a *Sophisticated* Sender sends message u with probability y, where $as_t/s_s < y < (1–a)s_t/s_s$. Each of these y values leads to u and d games with a different, unique mixed-strategy equilibrium. In these equilibria a *Sophisticated* Sender plays U with probability $1–a/(1+a)[ys_s/(s_t+ys_s)] = [1– as_t/ys_s]/(1+a)$ in the u game and $1–a/(1+a)[(1–y)s_s/\{s_l + (1–y)s_s\}] = [1– as_l/(1–y)s_s]/(1+a)$ in the d game; a *Sophisticated* Receiver plays L with probability $[1– (1+a)r_b]/(1+a)r_s$ in the u game and $[1– (1+a)r_b]/(1+a)r_s$ in the d game; a *Sophisticated* Sender's equilibrium expected payoff is $a/(1+a)$; and a *Sophisticated* Receiver's equilibrium expected payoff is $–a/(1+a)$.[21]

[20]m refers to a probability mixture over messages u and d, and M_u (M_d) refers to the player's part of the relevant mixed-strategy equilibrium in the u (d) game, both described precisely in Proposition 1.
[21]Thus, there are multiple sequential equilibria, but sequential equilibrium is generically essentially unique in that in each role, they all have the same expected payoffs for *Sophisticated* and *Mortal* players.

In case (i), if $r_b > 1/(1+a)$, there is a unique mixed-strategy sequential equilibrium, in which a *Sophisticated* Sender sends message u with probability $y = s_t/as_s$ and plays D in the u game and U in the d game; a *Sophisticated* Receiver plays R in the u game and the d game; a *Sophisticated* Sender's expected payoff is $(s_t/as_s)(r_b + r_s) + (1- s_t/as_s)ar_b$, and a *Sophisticated* Receiver's expected payoff is $-s_t/[a(1+a)s_s]$.

In case (ii), where $r_i > r_b$, $ar_i + r_b < 1$, and $s_s > as_l$, the conclusions are the same as in case (i), but with the roles of r_i and r_b, and of s_l and s_t, reversed.

It may seem surprising that a *Sophisticated* Receiver's strategy is R,R in *all* pure-strategy sequential equilibria. This conclusion's asymmetry across actions stems from the fact that because $a > 1$, all *Mortal* Senders play U, and it holds trivially if there are enough *Mortal* Senders to make R a dominant strategy in the underlying game; but the conclusion holds even if there are not enough *Mortal* Senders, as long as the game has a pure-strategy sequential equilibrium. The reason is that if a *Sophisticated* Sender deviates from his pure-strategy equilibrium message, the deviation "proves" to a *Sophisticated* Receiver that the Sender is *Mortal,* making his best response R off the equilibrium path. But in the only pure-strategy equilibria (sequential or not) in which a *Sophisticated* Receiver's strategy is *not* R,R, a *Sophisticated* Sender plays U on the equilibrium path, so a *Sophisticated* Receiver must also play R on the equilibrium path.

The rest of Proposition 1's conclusions concerning pure-strategy equilibria are straightforward, given that a *Sophisticated* Receiver always plays R,R. Because a *Sophisticated* Sender cannot truly fool a *Sophisticated* Receiver in equilibrium, whichever action he chooses in the underlying game, it is always best to send the message that fools whichever type of *Mortal* Receiver, *Believer* or *Inverter*, is more likely. The only remaining choice is whether to play U or D, when, with the optimal message, the former action fools $\max\{r_b,r_i\}$ *Mortal* Receivers at a gain of a per unit and the latter fools them at a gain of 1 per unit, but also "fools" r_s *Sophisticated* Receivers. Simple algebra reduces this question to whether $a \max\{r_b,r_i\} + \min\{r_b,r_i\} > 1$ or < 1.

It is clear from Figure 4 that the model's pure-strategy sequential equilibria avoid the perverse comparative statics of equilibrium mixed strategies with respect to a in the standard analysis, noted in fn. 9. Within the region that supports a given pure-strategy

equilibrium, *a* does not affect *Mortal* or *Sophisticated* players' strategies at all. However, as intuition suggests, increasing *a* always enlarges the set of type frequencies that support equilibria in which a *Sophisticated* Sender's equilibrium action is U (B, B', E, or E').

Proposition 1's conclusions concerning mixed-strategy equilibria in case (i) if $r_b < 1/(1+a)$ (Γ'_m), or in case (ii) if $r_i < 1/(1+a)$ (Z'_m), are straightforward extensions of the standard analysis to parameter configurations in which the probabilities of a *Sophisticated* Sender and Receiver are both high.[22] But in case (i) if $r_b > 1/(1+a)$ (Γ_m), or case (ii) if $r_i > 1/(1+a)$ (Z_m), the model has unique mixed-strategy sequential equilibria with a different character, in which randomization is confined to a *Sophisticated* Sender's message, and serves to "punish" a *Sophisticated* Receiver for deviating from R,R in a way that relaxes the $s_s \leq as_t$ or $s_s \leq as_l$ constraint whose violation prevents a *Sophisticated* Sender from realizing the higher expected payoff of equilibrium Γ or Z. These equilibria are otherwise similar to the pure-strategy equilibria Γ or Z for adjoining parameter configurations, and converge to them as the relevant population parameters converge.

In both kinds of mixed-strategy equilibrium, players' strategies are determined by simple, portable behavioral parameters as for pure-strategy equilibria; but both share some of the delicacy of HM's equilibria, and of mixed-strategy equilibria more generally.

To assess the model's ability to explain Operation Fortitude, consider the parameter values that lead to the sequential equilibria Γ or Γ'. In general, the conditions for Γ or Γ' to be a sequential equilibrium are $r_b > r_i$, $ar_b + r_i < 1$, and $s_s < as_t$. $r_b > r_i$ reflects a preponderance of *Believers* over *Inverters* that seems quite plausible, so I shall assume it.[23] Given this, suppose that $r_b = cr_i$ and $s_l = cs_t$ for some constant *c*. Then, Γ or Γ' is sequential iff $r_b < c/((ac+1)$ and $s_s < a/(1+a+c)$. When $a = 1.4$, as in Figure 4, and $c = 3$, which seem plausible values, these conditions reduce to $r_b < 0.58$ and $s_s < 0.26$, plausible ranges for these parameters.[24] By contrast, the conditions for the "reverse Fortitude" sequential equilibria E or E' are $r_b > r_i$ and $ar_b + r_i > 1$. Again assuming that $r_b > r_i$ and $r_b = cr_i$, E or E' is sequential iff $r_b > c/((ac+1)$. When $a = 1.4$ and $c = 3$, the condition reduces to $r_b > 0.58$. Thus, if the Germans are thought sufficiently likely to be

[22]This part of Proposition 1 shows that the standard analysis is robust to some bounded rationality of the kind considered here, but this may be an artifact of the discrete action spaces of the underlying game.
[23]If $r_i > r_b$ the sequential equilibria Z or Z' would duplicate the outcomes of Γ or Γ', with inverted messages.
[24]Higher values of *a* make the first condition more stringent and the second less stringent.

gullible, it is better to feint at Normandy and attack at Calais. However, in this application $r_b > 0.58$ seems less realistic than the conjunction of $r_b < 0.58$ and $s_s < 0.26$.[25]

III. Welfare Analysis

This section conducts a welfare analysis of the model's sequential equilibria, comparing the expected payoffs of *Mortal* and *Sophisticated* types. The comparisons use actual rather than anticipated expected payoffs for *Mortal* types, whose beliefs may be incorrect. I focus on cases where $r_b > r_i$; transposition yields the results when $r_i > r_b$.

Sender type	E or E' equilibrium message, action, and payoff	Γ or Γ' equilibrium message, action, and payoff	Γ_m equilibrium message, action(s), and payoff	Γ'_m equilibrium message, action(s), and payoff
Liar	d, U, ar_b	d, U, ar_b	d, U, ar_b	d, U, $a/(1+a)$
Truthteller	u, U, ar_i	u, U, ar_i	u, U, ar_i	u, U, $a/(1+a)$
Sophisticated	d, U, ar_b	u, D, $r_b + r_s$	m, D|u, U|d, (s_t/as_s)x $(r_b+r_s)+(1-s_t/as_s)ar_b$	m, M_u|u M_d|d, $a/(1+a)$
Receiver type	**E or E' equilibrium action|u, action|d, and payoff**	**Γ or Γ' equilibrium action|u, action|d, and payoff**	**Γ_m equilibrium action|u, action|d, and payoff**	**Γ'_m equilibrium action|u, action|d, and payoff**
Believer	R, L, $-a(s_l + s_s)$	R, L, $-as_l - s_s$	R, L, $-as_l - s_s[(s_t/as_s) + (1- s_t/as_s)a] = -a(s_l+s_s) - s_t/a + s_t$	R, L, $-a/(1+a)$
Inverter	L, R, $-as_t$	L, R, $-as_t$	L, R, $-as_t$	L, R, $-a/(1+a)$
Sophisticated	R, R, 0	R, R, $-s_s$	$-s_s(s_t/as_s) = -s_t/a$	M_u,M_d, $-a/(1+a)$

Table 3. Expected payoffs of *Mortal* and *Sophisticated* Sender and Receiver types ($r_b > r_i$)

Table 3 lists all types' messages, actions, and expected payoffs on the possible sequential equilibrium paths, extending Figure 2's payoff calculations and Proposition 1's

[25]The plausibility of the Γ or Γ' equilibria may be further enhanced by the human tendency to overrate one's own strategic sophistication relative to others'. Further, Γ_m is behaviorally similar to Γ, and so one might relax Γ's restriction on s_s, at the cost of a random prediction of a *Sophisticated* Sender's message.

characterization of equilibrium behavior from *Sophisticated* to *Mortal* players.[26] The table shows that *Sophisticated* players in either role have expected payoffs at least as high as their *Mortal* counterparts'. This much is true by definition, because *Sophisticated* players can always mimic *Mortal* players, but in pure-strategy equilibria, *Sophisticated* players have strictly higher payoffs. *Sophisticated* Senders' advantage over *Mortal* Senders in these equilibria stems from their ability to avoid being fooled and to choose which type(s) to fool. *Sophisticated* Receivers' advantage comes from their ability to avoid being fooled, or to choose the least costly way to be "fooled."

Sophisticated players enjoy a smaller advantage in the mixed-strategy sequential equilibria Γ_m or Z_m, but for similar reasons. By contrast, in the mixed-strategy sequential equilibria Γ_m' or Z_m', *Sophisticated* players' equilibrium mixed strategies completely offset each other's gains from fooling *Mortal* Receivers, and in each role, *Sophisticated* and *Mortal* players have the same expected payoffs. Thus, in this case, the prevalence of *Sophisticated* players protects *Mortal* players from exploitation.

IV. Related Work

This section briefly discusses related work.

Sobel (1985) was the first to propose an equilibrium explanation of lying, studying an "enemy" Sender's incentives in repeated interaction to build and eventually exploit a reputation for being a "friend" of the Receiver's. His analysis focused on communication of private information in settings with asymmetric information about the Sender's preferences, as opposed to the asymmetric information about the Sender's and the Receiver's strategic thinking analyzed here. Benabou and Laroque (1992) extended Sobel's analysis to allow the Sender to have noisy private information, and used it to analyze the use of inside information to manipulate financial markets.[27]

Farrell and Gibbons' (1989) analysis of costless communication to multiple audiences has already been mentioned. Loury (1994) provides a different perspective on the issues that arise with multiple audiences.

---

[26]Table 3 sometimes combines the equilibrium-path outcomes of more than one equilibrium, to save space. The equilibrium actions may therefore differ from *Sophisticated* players' sequential equilibrium strategies.
[27]See also John Morgan and Philip Stocken's (2001) analysis of financial analysts' incentives to reveal private information to investors, and some of the references cited there.

Finally, Conlisk (2001) studied the adaptive dynamics of selection in favor of types with higher payoffs among a different set of types, *Trickster*, *Avoider*, and *Sucker*, taking the "fooling technology" as given (fn. 18). He showed that if those types have successively lower costs they can coexist in long-run equilibrium, proving (in a special case) P. T. Barnum's dictum, "There's a sucker born every minute, and two to take him."

Section III's calculation of the types' equilibrium expected payoffs in the present model could be used to conduct a similar adaptive analysis. I conjecture that unless *Sophisticated* players have higher costs, their payoff advantage in pure-strategy (and some mixed-strategy) equilibria will lead their relative frequencies to grow until the population frequencies enter the region of mixed-strategy equilibria in which all types' expected payoffs are equal (region Γ'-Z' in Figure 4). The population can then be expected to drift among a continuum of neutral steady states in that region. If *Sophisticated* players have slightly higher costs, the population frequencies should approach and remain near the boundary of region Γ'-Z', without entering it.[28] This would also allow *Sophisticated* and *Mortal* players to coexist in long-run equilibrium, justifying the assumptions about the type frequencies maintained here.

V. Conclusion

In this paper, I have proposed a way to model active misrepresentation of intentions to competitors or enemies. The model focuses on the strategic interaction between rational and boundedly rational types of players in a game of conflicting interests, with one-sided preplay communication via costless messages.

Allowing for the possibility of bounded rationality yields a sensible account of lying via costless, noiseless messages, and simplifies many aspects of the analysis of games with communication. For many parameter configurations, in contrast to a standard analysis of communication with conflicting interests, the model has generically unique pure-strategy sequential equilibria, in which rational players exploit boundedly rational

[28]This conclusion is not immediate because the present model has two player populations and a more complex pattern of payoff advantages than Conlisk's model. There, taking cost differences into account, in pairwise interactions *Tricksters* do better than *Suckers*, *Suckers* do better than *Avoiders*, and *Avoiders* do better than *Tricksters*. Here, with equal costs for *Mortal* players and higher costs for *Sophisticated* players in each player role, *Liars* do better than *Truthtellers* iff there are more *Believers* than *Inverters*, and

players, but are not themselves fooled. In these equilibria, players' strategies are determined by simple, portable behavioral parameters, and can be identified by iterated elimination of conditionally dominated strategies. One of these equilibria plausibly explains the Allies' decision to feint at Calais and attack at Normandy on D-Day.

For other parameter configurations, the model has generically unique or essentially unique mixed-strategy sequential equilibria, in which rational players' equilibrium strategies offset each other's gains from fooling boundedly rational players, completely protecting them from exploitation. In these equilibria, players' strategies are still determined by simple, portable behavioral parameters, but they share some of the delicacy of mixed-strategy equilibria in other games.

I hope that the methods for modeling bounded strategic rationality used here can elucidate strategic communication when players' interests are not in conflict, and can also be used to create behaviorally realistic models of strategic behavior in other applications.

References

Aumann, Robert, and Adam Brandenburger (1995): "Epistemic Conditions for Nash Equilibrium," *Econometrica*, 63, 1161-1180.

Benabou, Roland, and Guy Laroque (1992): "Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility," *Quarterly Journal of Economics*, 107, 921-958.

Blume, Andreas, Douglas DeJong, Yong-Gwan Kim, and Geoffrey Sprinkle (1998): "Experimental Evidence on the Evolution of Meaning of Messages in Sender-Receiver Games," *American Economic Review*, 88, 1323-1340.

Blume, Andreas, Douglas V. DeJong, Yong-Gwan Kim, and Geoffrey Sprinkle (2001): "Evolution of Communication with Partial Common Interest," *Games and Economic Behavior*, ?, ???-???.

Camerer, Colin, Teck-Hua Ho, and Juin-Kuan Chong (2002): "Sophisticated EWA Learning and Strategic Teaching in Repeated Games, *Journal of Economic Theory* ??, ???-???.

Believers do better than *Inverters* iff there are more *Truthtellers* than *Liars*. *Sophisticated* players may do better or worse than their *Mortal* counterparts, depending on the parameter configuration and the costs.

Conlisk, John (2001): "Costly Predation and the Distribution of Competence," *American Economic Review*, 91, 475-484.

Costa-Gomes, Miguel (2002): "A Suggested Interpretation of Some Experimental Results on Preplay Communication," *Journal of Economic Theory*, ??, ???-???.

Costa-Gomes, Miguel, Vincent P. Crawford, and Bruno Broseta (2001): "Cognition and Behavior in Normal-Form Games: an Experimental Study," *Econometrica*, 69, 1193-1235.

Crawford, Vincent P. (1998): "A Survey of Experiments on Communication via Cheap Talk," *Journal of Economic Theory*, 78, 286-298.

Crawford, Vincent P., and Dennis E. Smallwood (1984): "Comparative Statics of Mixed-Strategy Equilibria in Non-Cooperative Two-Person Games," *Theory and Decision*, 61, 225-232.

Crawford, Vincent P., and Joel Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50, 1431-1451.

Farrell, Joseph (1988): "Communication, Coordination and Nash Equilibrium," *Economics Letters*, 27, 209-214; see also (1990) "Erratum," 33, 299.

Farrell, Joseph (1993): "Meaning and Credibility in Cheap-Talk Games," *Games and Economic Behavior*, 5, 514-531.

Farrell, Joseph, and Robert Gibbons (1989): "Cheap Talk with Two Audiences," *American Economic Review*, 79, 1214-1223.

Farrell, Joseph, and Matthew Rabin (1996): "Cheap Talk," *Journal of Economic Perspectives*, 10, 103-118.

Harrison, Gordon (1951): *Cross-Channel Attack*, Washington, D.C.: Department of the Army, Center of Military History Publication 7-4, available online at http://www.army.mil/cmh-pg/books/wwii/7-4/7-4_a.HTM

Hendricks, Kenneth, and R. Preston McAfee (2000): "Feints," manuscript, University of Texas.

Kemp, Anthony (1994): *D-Day and the Invasion of Normandy*. New York: Harry N. Abrams.

Krugman, Paul (2001): "Reckonings: Truth and Lies," *New York Times*, August 28, 2001.

Loury, Glenn (1994): "Self-Censorship in Public Discourse: A Theory of 'Political

Correctness' and Related Phenomena," *Rationality and Society*, 6, 428-461.

Morgan, John, and Phillip Stocken (2001): "An Analysis of Stock Recommendations," manuscript, Princeton University.

Rabin, Matthew (1994): "A Model of Pre-Game Communication," *Journal of Economic Theory*, 63, 370-391.

Reny, Philip (1992): "Backward Induction, Normal-Form Perfection and Explicable Equilibria," *Econometrica*, 60, 627-649.

Mike Royko, "Read my Lips: Taxes Going Up," *Chicago Tribune*, November 8, 1988, available online at http://www.resnet.trinity.edu/mshay2/royko/881108.htm.

Selten, Reinhard (1998): "Features of Experimentally Observed Bounded Rationality," *European Economic Review*, 42, 413-436.

Shimoji, Makoto, and Joel Watson (1998): "Conditional Dominance, Rationalizability, and Game Forms," *Journal of Economic Theory*, 83, 161-195.

Sobel, Joel (1985): "A Theory of Credibility," *Review of Economic Studies*, 52, 557-573.

Stahl, Dale, and Paul Wilson (1995): "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10, 218-254.

Von Neumann, John, and Oskar Morgenstern (1953): *Theory of Games and Economic Behavior* (3rd edition; first edition 1944), New York: John Wiley and Sons.

Appendix. Proof of Proposition 1

I begin by characterizing the equilibria of the u and d games (Figure 3), as determined by a *Sophisticated* Receiver's belief, z, that the Sender is *Sophisticated*.

LEMMA 1. The u game has a generically unique equilibrium as follows:

(i) U,R is a pure-strategy equilibrium iff $r_i > 1/(1+a)$;

(ii) D,L is a pure-strategy equilibrium iff $r_b > a/(1+a)$ and $z > a/(1+a)$;

(iii) D,R is a pure-strategy equilibrium iff $r_i < 1/(1+a)$ and $z < a/(1+a)$; and

(iv) there is a mixed-strategy equilibrium, with Pr{*Sophisticated* Sender plays U} = $1 - a/(1+a)z$, Pr{*Sophisticated* Receiver plays L} = $[1- (1+a)r_i]/(1+a)r_s$, *Sophisticated*

Sender's expected payoff $a/(1+a)$, and *Sophisticated* Receiver's expected payoff $-a/(1+a)$, iff $r_i < 1/(1+a)$, $r_b < a/(1+a)$ and $z > a/(1+a)$.

The d game has a generically unique equilibrium as follows:[29]

(i) U,R is a pure-strategy equilibrium iff $r_b > 1/(1+a)$;

(ii) D,L is a pure-strategy equilibrium iff $r_i > a/(1+a)$ and $z > a/(1+a)$;

(iii) D,R is a pure-strategy equilibrium iff $r_b < 1/(1+a)$ and $z < a/(1+a)$; and

(iv) there is a mixed-strategy equilibrium, with Pr{*Sophisticated* Sender plays U} = $1 - a/(1+a)z$, Pr{*Sophisticated* Receiver plays L} = $[1-(1+a)r_b]/(1+a)r_s$, *Sophisticated* Sender's expected payoff $a/(1+a)$, and *Sophisticated* Receiver's expected payoff $-a/(1+a)$, iff $r_b < 1/(1+a)$, $r_i < a/(1+a)$ and $z > a/(1+a)$.

PROOF: Straightforward calculations, noting that (U,L) is never an equilibrium, and, because $r_i > 1/(1+a)$ and $r_b > a/(1+a)$ or vice versa are inconsistent, the conditions for (i)-(iv) are mutually exclusive and (with nongeneric exceptions) collectively exhaustive. □

Lemmas 2-3, which correspond to the pure- and mixed-strategy cases considered in Proposition 1, characterize the sequential equilibria of the reduced game.

LEMMA 2: Unless either $r_b > r_i$, $ar_b + r_i < 1$, and $s_s > as_t$, or $r_i > r_b$, $ar_i + r_b < 1$, and $s_s > as_l$, the reduced game has a generically unique sequential equilibrium in pure strategies, in which a *Sophisticated* Sender's and Receiver's strategies are as given in Table 2 and Figure 4. In these sequential equilibria, a *Sophisticated* Receiver's strategy is R,R; and a *Sophisticated* Sender plays U (D) on the equilibrium path iff $a \max\{r_b,r_i\} + \min\{r_b,r_i\} >$ (<) 1 and sends message d (u) iff $r_b >$ (<) r_i. *Sophisticated* players' sequential equilibrium strategies are their unique extensive-form rationalizable strategies, identifiable by at most three steps of iterated conditional dominance.

PROOF: Because all types have positive prior probability and *Liars* and *Truthtellers* send different messages, all messages have positive probability in equilibrium. Further, in any pure-strategy sequential equilibrium, a *Sophisticated* Sender's message is pooled with

[29]The characterization here is identical to that for the "u" subgame, with the roles of r_b and r_i interchanged.

either *Liars*' or *Truthtellers*' message, so a deviation to the other message makes $z = 0$. In the u or d game that follows such a deviation, R is a conditionally dominant strategy for a *Sophisticated* Receiver; and a *Sophisticated* Sender's unique best response is U (D) iff $r_i >$ (<) $1/(1+a)$ in the u game and U (D) iff $r_b >$ (<) $1/(1+a)$ in the d game by Lemma 1.

All that remains is to identify the strategy combinations in Figure 2 that are equilibria for some parameter configurations, use these conditions to check which configurations make them sequential, and check the other conclusions of the lemma.

Identifying the configurations by the Greek capital letters in Figure 2, Δ and Δ' are equilibria iff $r_b > ½$. For Δ to be sequential, U,L must be an equilibrium in the u game when $z = 0$, which is never true. For Δ' to be sequential, D,L must be an equilibrium in the u game when $z = 0$, which is never true. Thus neither Δ nor Δ' is ever sequential. Similarly, A and A' are equilibria iff $r_b > ½$, but neither A nor A' is ever sequential.

E and E' are equilibria iff $r_b > r_i$ and $ar_b > r_b + r_s$, which reduces to $ar_b + r_i > 1$. For E to be sequential, U,R must be an equilibrium in the u game when $z = 0$, which is true iff $r_i > 1/(1+a)$. Thus E is sequential iff $r_b > r_i$, $ar_b + r_i > 1$, and $r_i > 1/(1+a)$, where the second condition is implied by the first and third. For E' to be sequential, D,R must be an equilibrium in the u game when $z = 0$, which is true iff $r_i < 1/(1+a)$. Thus E' is sequential iff $r_b > r_i$, $ar_b + r_i > 1$, and $r_i < 1/(1+a)$. Similarly, B and B' are equilibria iff $r_i > r_b$ and $ar_i + r_b > 1$; B is sequential iff $r_i > r_b$, $ar_i + r_b > 1$, and $r_b > 1/(1+a)$, where the second condition is implied by the first and third; and B' is sequential iff $r_i > r_b$, $ar_i + r_b > 1$, and $r_b < 1/(1+a)$.

Γ and Γ' are equilibria iff $s_s < as_t$, $r_b > r_i$, and $r_b + r_s > ar_b$, which reduces to $ar_b + r_i < 1$. For Γ to be sequential, U,R must be an equilibrium in the d game when $z = 0$, which is true iff $r_b > 1/(1+a)$. Thus Γ is sequential iff $s_s < as_t$, $r_b > r_i$, $ar_b + r_i < 1$, and $r_b > 1/(1+a)$. For Γ' to be sequential, D,R must be an equilibrium in the d game when $z = 0$, which is true iff $r_b < 1/(1+a)$. Thus Γ' is sequential iff $s_s < as_t$, $r_b > r_i$, $ar_b + r_i < 1$, and $r_b < 1/(1+a)$, where the second condition is implied by the first and third. Similarly, Z and Z' are equilibria iff $s_s < as_l$, $r_i > r_b$, and $ar_i + r_b < 1$; Z is sequential iff $s_s < as_l$, $r_i > r_b$, $ar_i + r_b < 1$, where the second condition is implied by the first and third; and $r_i > 1/(1+a)$. and Z' is sequential iff $s_s < as_l$, $r_i > r_b$, $ar_i + r_b < 1$, and $r_i < 1/(1+a)$.

In each case, the generic uniqueness of *Sophisticated* players' best responses can be verified by iterated conditional dominance, starting with the pure-strategy equilibria in the 2x2 u and d games. The remaining conclusions are easily verified by inspection. □

LEMMA 3: If either (i) $r_b > r_i$, $ar_b + r_i < 1$, and $s_s > as_t$; or (ii) $r_i > r_b$, $ar_i + r_b < 1$, and $s_s > as_l$, the reduced game has a generically unique or essentially unique mixed-strategy sequential equilibrium, in which a *Sophisticated* Sender's and Receiver's strategies are as given in Table 2 and Figure 4. In case (i), if $r_b < 1/(1+a)$, there are multiple mixed-strategy sequential equilibria, in each of which a *Sophisticated* Sender sends message u with probability y, where $as_t/s_s < y < (1-a)s_t/s_s$. Each of these y values leads to u and d games with a different, unique mixed-strategy equilibrium. In these equilibria a *Sophisticated* Sender plays U with probability $1-a/(1+a)[ys_s/(s_t+ys_s)] = [1-as_t/ys_s]/(1+a)$ in the u game and $1-a/(1+a)[(1-y)s_s/\{s_l + (1-y)s_s\}] = [1- as_l/(1-y)s_s]/(1+a)$ in the d game; a *Sophisticated* Receiver plays L with probability $[1- (1+a)r_b]/(1+a)r_s$ in the u game and $[1- (1+a)r_b]/(1+a)r_s$ in the d game; a *Sophisticated* Sender's equilibrium expected payoff is $a/(1+a)$; and a *Sophisticated* Receiver's equilibrium expected payoff is $-a/(1+a)$.

In case (i), if $r_b > 1/(1+a)$, there is a unique mixed-strategy sequential equilibrium, in which a *Sophisticated* Sender sends message u with probability $y = s_t/as_s$ and plays D in the u game and U in the d game; a *Sophisticated* Receiver plays R in the u game and the d game; a *Sophisticated* Sender's expected payoff is $(s_t/as_s)(r_b + r_s) + (1- s_t/as_s)ar_b$, and a *Sophisticated* Receiver's expected payoff is $-s_t/[a(1+a)s_s]$.

In case (ii), where $r_i > r_b$, $ar_i + r_b < 1$, and $s_s > as_l$, the conclusions are the same as in case (i), but with the roles of r_i and r_b, and of s_l and s_t, reversed.

PROOF: In case (i), if $r_b < 1/(1+a)$, and if $z < a/(1+a)$ in either the u or the d game, D,R would be its unique equilibrium. But then, in this case, a *Sophisticated* Sender would prefer to send the message that led to that game with probability 1, and with $z = 0$, players would also have pure best responses in the other game by Lemma 1. But the proof of Lemma 2 shows that there are no pure-strategy sequential equilibria in this case when $s_s > as_t$. If, instead, $z > a/(1+a)$ in each game, in this case the u and d games have unique mixed-strategy equilibria as characterized in Lemma 1. $ys_s/(s_t+ys_s) > a/(1+a)$ and (1–

$y)s_s/[s_l + (1–y)s_s] > a/(1+a)$ provided that $as_t/s_s < y < 1–as_l/s_s$, which is always feasible when $s_s > as_t$. Because a *Sophisticated* Sender's expected payoff is $a/(1+a)$ in either the u or the d game, he is willing to randomize with any such y. The rest of the proof in this case is a straightforward translation of the conclusions of Lemma 1.

In case (i), if $r_b > 1/(1+a)$, the d game always has a unique pure-strategy equilibrium U,R, with expected payoff ar_b for a *Sophisticated* Sender. If $y = 0$, the u game would be off the equilibrium path, so message u would make $z = 0$, and the u game would have a unique pure-strategy equilibrium D,R, with payoff $r_b + r_s > ar_b$ for a *Sophisticated* Sender. Thus there cannot be an equilibrium in this case with $y = 0$. Similarly, if $y = 1$, iff $r_b < a/(1+a)$ the u game has a unique mixed-strategy equilibrium, with $z = s_s/(s_t+s_s) > a/(1+a)$ and expected payoff $a/(1+a) < ar_b$ for a *Sophisticated* Sender. If $y = 1$ and $r_b > a/(1+a)$, the u game has a unique pure-strategy equilibrium, D,L, and expected payoff $r_b < ar_b$ for a *Sophisticated* Sender. Thus there cannot be an equilibrium with $y = 1$. Because $r_b + r_s > ar_b > a/(1+a)$ in this case, a *Sophisticated* Sender's optimal choice of y maximizes $y(r_b + r_s) + (1– y)ar_b$ subject to the constraint that D,R is an equilibrium in the u game, which is true in this case iff $z = ys_s/(s_t+ys_s) \leq a/(1+a)$, or equivalently $y \leq s_t/as_s$. Thus, a *Sophisticated* Sender's optimal message strategy is $y = s_t/as_s$.[30] The rest of the proof follows directly from Lemma 1, noting that a *Sophisticated* Receiver's expected payoff is $–(s_t/as_s)\,[(s_t/as_s)s_s]/[s_t+(s_t/as_s)s_s] = –s_t/[a(1+a)s_s]$. □

Lemma 3 completes the proof of Proposition 1.

[30] A *Sophisticated* Sender's best response is not $y = 1$ in this case, because in the reduced game *Mortal* players' responses are treated as part of the payoff function, which effectively constrain his choice of y.

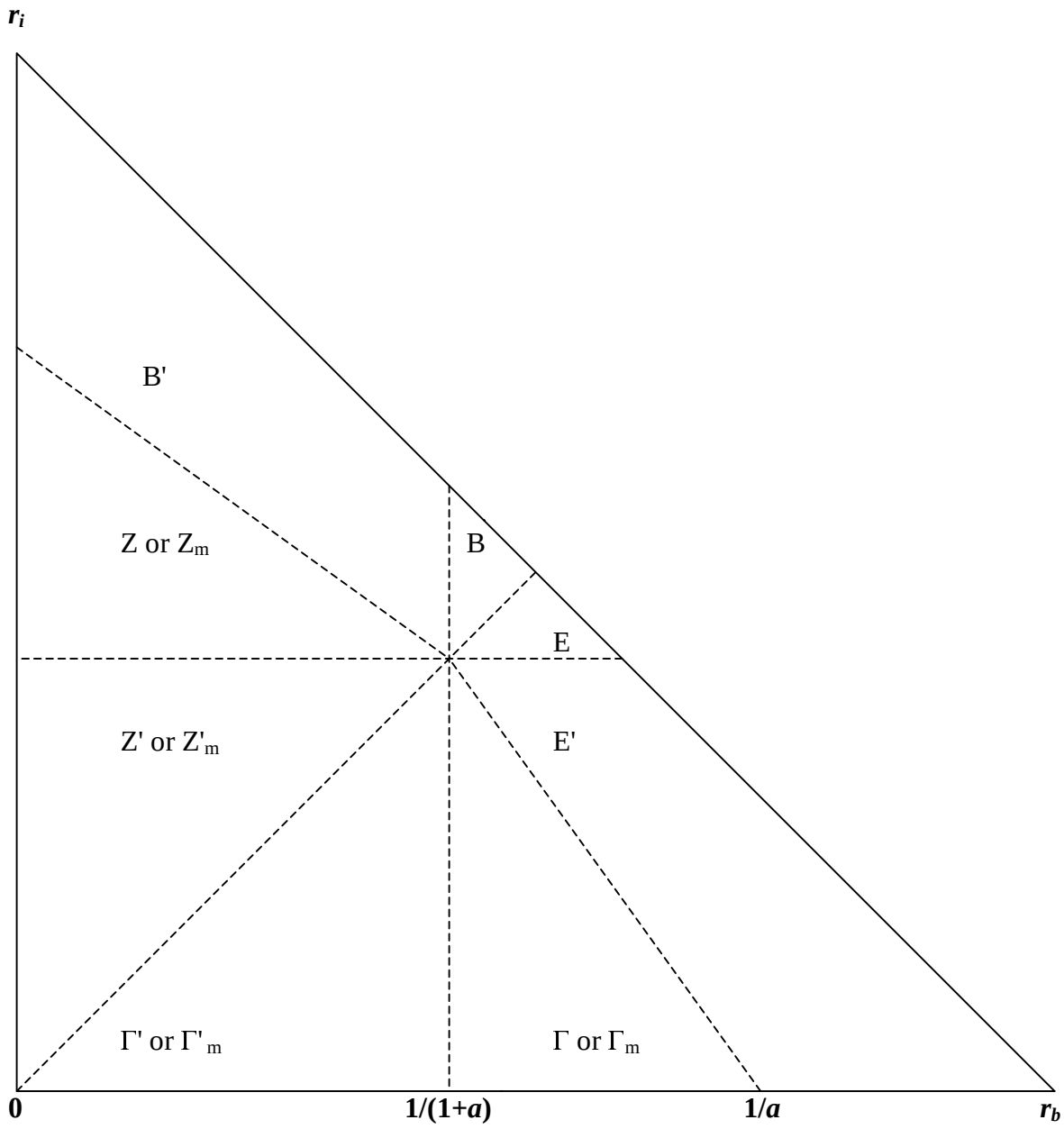


Figure 4. Sequential equilibria when $a = 1.4$
(subscript m denotes sequential equilibria when $s_s > as_t$ (as_l) in Γ or Γ' (Z or Z')