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

Vince Crawford, UCSD, October 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."—Lincoln

"Now give Barnum his due."—John Conlisk

Introduction

Lying to competitors, enemies, even friends with different preferences is an important phenomenon, but hard to explain using standard game-theoretic methods, which assume rational expectations

Focus here on active misrepresentation rather than less-than-full disclosure, and on signaling intentions rather than private information

Consider a simple model in which Sender sends Receiver a costless message, u or d, about intended action in zero-sum two-person game

Sender and Receiver then choose actions simultaneously; the structure is common knowledge

		Receiver: Left	Receiver: Right
Sender	Up	$-a$ $a > 1$	0 0
	Down	0 0	-1 1

Figure 1. The underlying game

In a standard equilibrium analysis, the Sender's message is uninformative and the Receiver ignores it; underlying game is then played according to its unique mixed-strategy equilibrium: U with probability 1/(1+*a*), L with probability 1/(1+*a*), with Sender's and Receiver's expected payoffs *a*/(1+*a*) and –*a*/(1+*a*)

Sender's message is uninformative, but no one is fooled by it; need something different to understand active misrepresentation

Possible escapes:

- Private information about preferences: Sobel's (1985) analysis of an "enemy" Sender's incentives in repeated interaction to build and eventually exploit a reputation for being a "friend" of the Receiver's

- Costly, noisy messages: Hendricks and McAfee's ("HM's") analysis of Operation Fortitude, the Allies' misrepresentation of intention to attack at Normandy rather than Calais on D-Day:

Attacker chooses (possibly randomly) between two possible locations and allocates a fixed budget of force between them

Defender then privately observes a binary signal whose probability distribution depends on the attacker's allocation and allocates (possibly randomly) his own budget of force between locations

Attack location and force allocations determine zero-sum payoffs

Payoff function and signal distribution symmetric across locations

Equilibrium must involve some misrepresentation (attacker allocating force to both locations with positive probability), with some success (defender allocating force to both locations with positive probability)

When signal is not very informative, attacker allocates most force to one location but randomizes location, defender allocates entire force deterministically, to location more likely to be attacked

When signal is more informative, attacker randomizes allocation and location so signal is uninformative (with positive probability of assigning *less* force to attack location), defender randomizes

When signal is not very informative a reduction in noise hurts the attacker; but when it's more informative, a reduction benefits attacker

Problems (here and in other applications):

(i) Cost of faking is small, more like cheap talk than large allocations

(ii) Analysis ignores asymmetry between Normandy and Calais: Why not feint at Normandy and attack at Calais, particularly if the feint has a fair chance of success? Analysis shouldn't leave this to chance

(iii) Assumptions that rationality and beliefs are mutual knowledge are strained, especially in one-shot game when equilibria have delicate balance of mixed strategies depending on details of signal distribution

Model

My goal is to give a sensible account of active misrepresentation in a simpler game, with costless and noiseless messages

The key is allowing for the *possibility* of bounded strategic rationality

Reconsider the above game, for concreteness identifying the Sender with the Allies, U with attacking Calais, L with defending Normandy; $a > 1$ reflects greater ease of an unanticipated attack at Calais

Now, each player role is filled randomly from a separate distribution of decision rules, or *types*, with boundedly rational, or *Mortal*, types as well as to a fully strategically rational, or *Sophisticated*, type

Players don't observe others' types, but structure common knowledge

Sender's possible 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); Receiver's are (action|received u, action|received d) = (L,L), (L,R), (R,L), or (R,R)

Sender type	**Behavior (b.r. ≡ best response)**	**message, action|sent u, action|sent d**
Credible ≡	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 ≡	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. Plausible *Mortal* and *Sophisticated* Sender and Receiver types

Mortal types, like other boundedly rational types, use step-by-step procedures that generically determine unique, pure strategies, avoid simultaneous determination of the kind used to define equilibrium

Mortals' strategies determined independently of each other's and *Sophisticated* players' strategies, so can be treated as exogenous (though they affect others' payoffs); strategic analysis can focus on *reduced game* between possible *Sophisticated* players in each role

Reduced game is not zero-sum, messages are not cheap talk, and it has incomplete information; so analysis different, maybe more helpful

Observations

- *Wily* Sender, *Wj*, with *j* odd always lies; lump these *Mortal* Sender types together under the heading *Liars*

- *Wily* Sender with *j* even (including *Credible* as honorary *Wily* type, *W0*) always tells the truth; lump them together as *Truthtellers*

- *Skeptical* Receiver, *Sk*, with *k* odd always inverts the Sender's message, and with *k* even (including *Credulous* as *S0*) always believes it; lump them together as *Inverters and Believers*

- Behavior of Sender population can be summarized by $s_l \equiv$ Pr{Sender's a *Liar*}, $s_t \equiv$ Pr{Sender's a *Truthteller*}, and $s_s \equiv$ Pr{Sender's *Sophisticated*}, and behavior of Receiver population can be summarized by $r_i \equiv$ Pr{Receiver's an *Inverter*}, $r_b \equiv$ Pr{Receiver's a *Believer*}, and $r_s \equiv$ Pr{Receiver's *Sophisticated*}; assume these type probabilities are all strictly positive in both populations, and ignore nongeneric parameter configurations

- *Inverters* and *Believers* always choose different actions for a given message, but *Mortal* Sender types always play U on equilibrium path

- *Liars* therefore send message d and *Truthtellers* send message u; thus both messages have positive probability, and a *Sophisticated* Sender is always pooled with one *Mortal* Sender type

- After a message for which a *Sophisticated* Sender plays U with probability 1, a *Sophisticated* Receiver's best response is R

- Otherwise his best response may depend on his posterior *belief*, z, that Sender is *Sophisticated*: if x is message and y is *Sophisticated* Sender's probability of sending u, *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 reduced game between a *Sophisticated* Sender and Receiver
(Greek capitals identify pure-strategy equilibria (sequential or not) for some parameters)

		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

		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

Analysis

(E) d,U,U; R,R iff $r_b > r_i$, $ar_b + r_i > 1$, and $r_i > 1/(1+a)$ (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 $s_s < as_t$ (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 $r_b > 1/(1+a)$ (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 $s_s < as_l$ (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

μ When the probabilities of a *Sophisticated* Sender and Receiver are high, the reduced game has a generically essentially unique sequential equilibrium in mixed strategies; in this case *Sophisticated* players' equilibrium mixed strategies offset each other's gains from fooling *Mortal* players, *Sophisticated* players have the same expected payoffs as their *Mortal* counterparts, and all types' expected payoffs are the same as in the standard analysis

μ There are also hybrid mixed-strategy equilibria when a *Sophisticated* Sender (Receiver) has high (low) probabity, in which randomization is confined to the Sender's message, and "punishes" a *Sophisticated* Receiver for deviating from R,R in a way that allows the Sender to realize higher expected payoff; these equilibria are like the pure-strategy equilibria for adjoining parameter configurations, and converge to them as the relevant population parameters converge

μ When the probabilities of a *Sophisticated* Sender and Receiver are low, the reduced game has a generically unique sequential equilibrium in pure strategies, in which a *Sophisticated* Receiver can predict a *Sophisticated* Sender's action, and vice versa

μ *Sophisticated* Receiver's strategy is R,R in *all* pure-strategy sequential equilibria because if Sender deviates from his pure-strategy equilibrium message, it "proves" that Sender is *Mortal,* making Receiver's best response R; but in the only pure-strategy equilibria 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

μ 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

μ 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

Fortitude

μ When the probability of a *Sophisticated* Sender is low and the probability of a *Believer* is not too high, the model has a unique sequential equilibrium (Γ or Γ') 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 *Believer* play R—like defending Calais; *Sophisticated* Receiver plays R because being "fooled" at unit cost 1 by a *Sophisticated* Sender is preferable to being "fooled" at unit cost a by both kinds of *Mortal* Sender

μ Conditions for Γ or Γ' are $r_b > r_i$, $ar_b + r_i < 1$, and $s_s < as_t$; assume $r_b > r_i$, and suppose $r_b = cr_i$ and $s_l = cs_t$ for constant c; Γ or Γ' is sequential iff $r_b < c/((ac+1)$ and $s_s < a/(1+a+c)$; when $a = 1.4$ (Figure 4) and $c = 3$, these reduce to $r_b < 0.58$ and $s_s < 0.26$, plausible ranges

μ Conditions for "reverse Fortitude" equilibria E or E' are $r_b > r_i$ and $ar_b + r_i > 1$; if $r_b > r_i$ and $r_b = cr_i$, E or E' is sequential iff $r_b > c/((ac+1)$; when $a = 1.4$ and $c = 3$, this reduces to $r_b > 0.58$, maybe less realistic

μ In this explanation, players' sequential equilibrium strategies depend only on payoffs and population parameters that reflect simple, portable facts about behavior that could be learned in many imperfectly analogous conflict situations; in pure-strategy sequential equilibria, *Sophisticated* players' strategies are their unique extensive-form rationalizable strategies, identifiable by at most three steps of iterated conditional dominance (Shimoji-Watson (1998))

Welfare

Welfare analysis uses actual rather than anticipated expected payoffs for *Mortals*, whose beliefs may be incorrect; focus on $r_b > r_i$

μ *Sophisticated* players in either role have expected payoffs at least as high as their *Mortal* counterparts' by definition

μ In pure-strategy sequential equilibria, *Sophisticated* players in either role do strictly better than their *Mortal* counterparts; advantage comes from ability to avoid being fooled and/or choose which type(s) to fool

μ *Sophisticated* players enjoy a smaller advantage in the hybrid mixed-strategy sequential equilibria (Γ_m or Z_m), but for similar reasons

μ In mixed-strategy sequential equilibria that arise when probabilities of a *Sophisticated* Sender and Receiver are both high (Γ_m' or Z_m'), *Sophisticated* players' equilibrium mixed strategies offset each other's gains from fooling *Mortal* Receivers, and in each role *Sophisticated* and *Mortal* players have the same expected payoffs

μ This suggests that in an adaptive analysis of dynamics of type distribution, as in Conlisk (2001), frequencies of *Sophisticated* types will grow until the population is in or near (depending on costs) the region of mixed-strategy equilibria in which types' expected payoffs are equal (Γ'-*Z*' in Figure 4); this allows *Sophisticated* and *Mortal* players to coexist in long-run equilibrium, justifying assumptions

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 ($r_b > r_i$)

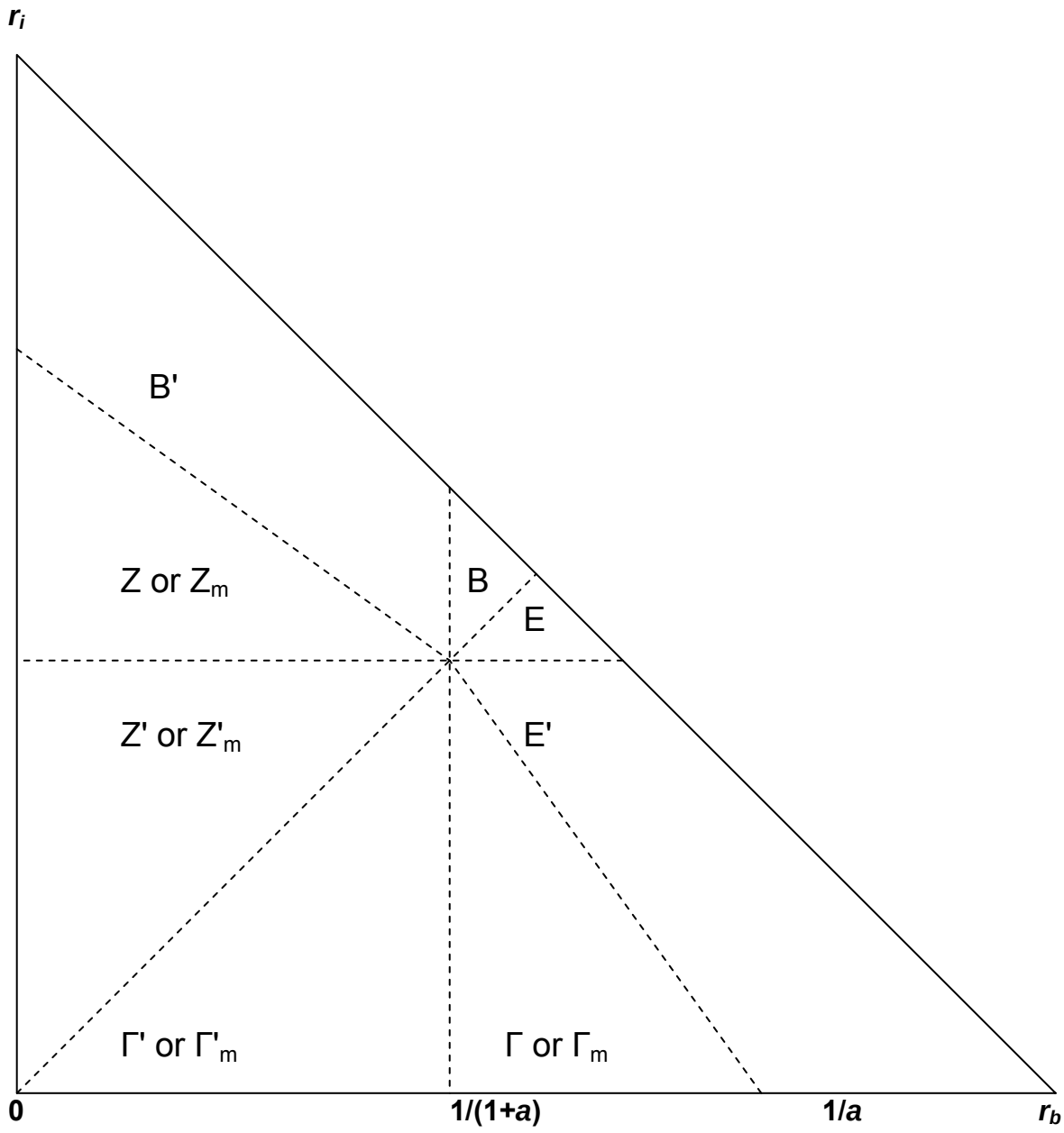

Figure 4. Sequential equilibria when a = 1.4
(subscript m denotes sequential equilibria when $s_s > as_t$ (as_l) in Γ or Γ' (Z or Z')