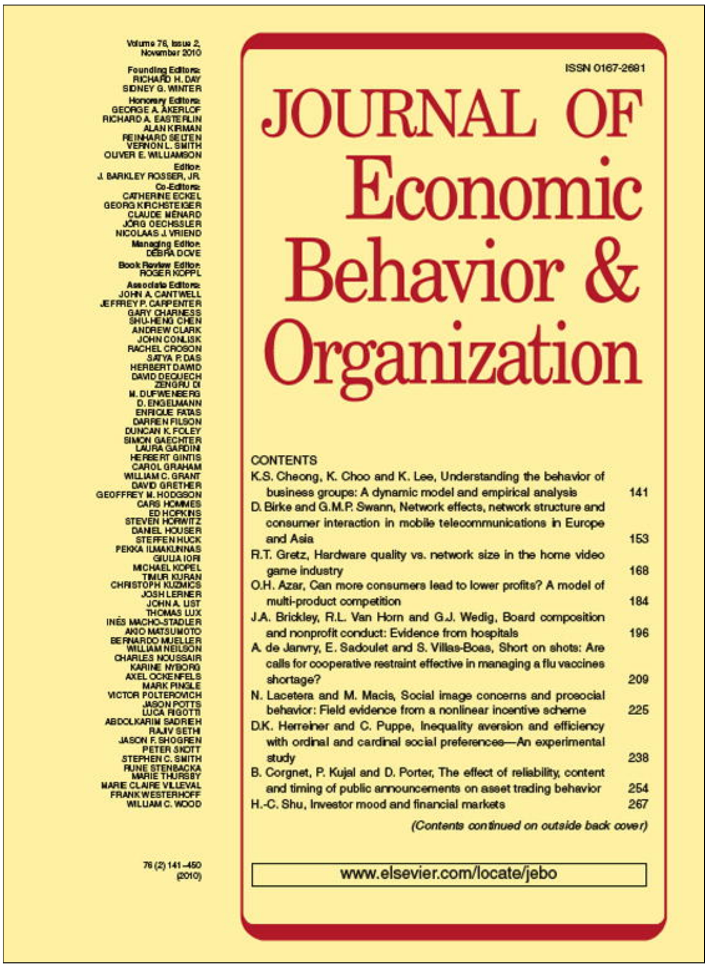
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



Volume 76, Issue 2,
November 2010

Founding Editors:
RICHARD H. DAY
SIDNEY G. WINTER
Honorary Editors:
GEORGE A. AKERLOF
RICHARD A. EASTERLIN
ALAN KIRMAN
REINHARD SELTEN
VERNON L. SMITH
OLIVER E. WILLIAMSON
Editor:
J. BARKLEY ROSSER, JR.
Co-Editors:
CATHERINE ECKEL
GEORG KIRCHSTEIGER
CLAUDE MÉNARD
JÖRG OECHSSLER
NICOLAAS J. VRIEND
Managing Editor:
DEBRA DOVE
Book Review Editor:
ROGER KOPPL
Associate Editors:
JOHN A. CANTWELL
JEFFREY P. CARPENTER
GARY CHARNESS
SHU-HENG CHEN
ANDREW CLARK
JOHN CONLISK
RACHEL CROSON
SATYA P. DAS
HERBERT DAWID
DAVID DEQUECH
ZENGRU DI
M. DUFWENBERG
D. ENGELMANN
ENRIQUE FATAS
DARREN FILSON
DUNCAN K. FOLEY
SIMON GAECHTER
LAURA GARDINI
HERBERT GINTIS
CAROL GRAHAM
WILLIAM C. GRANT
DAVID GRETHER
GEOFFREY M. HODGSON
CARS HOMMES
ED HOPKINS
STEVEN HORWITZ
DANIEL HOUSER
STEFFEN HUCK
PEKKA ILMAKUNNAS
GIULIA IORI
MICHAEL KOPEL
TIMUR KURAN
CHRISTOPH KUZMICS
JOSH LERNER
JOHN A. LIST
THOMAS LUX
INÉS MACHO-STADLER
AKIO MATSUMOTO
BERNARDO MUELLER
WILLIAM NEILSON
CHARLES NOUSSAIR
KARINE NYBORG
AXEL OCKENFELS
MARK PINGLE
VICTOR POLTEROVICH
JASON POTTS
LUCA RIGOTTI
ABDOLKARIM SADRIEH
RAJIV SETHI
JASON F. SHOGREN
PETER SKOTT
STEPHEN C. SMITH
RUNE STENBACKA
MARIE THURSBY
MARIE CLAIRE VILLEVAL
FRANK WESTERHOFF
WILLIAM C. WOOD

76 (2) 141–450
(2010)

ISSN 0167-2681

JOURNAL OF Economic Behavior & Organization

CONTENTS

(Contents continued on outside back cover)

www.elsevier.com/locate/jebo

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Journal of Economic Behavior & Organization 76 (2010) 318–324

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Truth and lie detection in bluffing☆

Håkan J. Holm

Lund University, Economics, P.O. Box 7082, 22007 Lund, Sweden

ARTICLE INFO

Article history:
Received 7 March 2007
Received in revised form 15 June 2010
Accepted 16 June 2010
Available online 23 June 2010

JEL classification:
C72
D82

Keywords:
Bluffing
Game theory
Truth detection
Lie detection

ABSTRACT

Beliefs in signals that reveal lies and truths are widespread. It is shown that such beliefs may be exploited strategically in signaling games of pure conflict of interest. Truth and lie detection is modeled by signals perceived by the receiver that are emitted with a probability contingent on the truth value of the sender's message. Truth or lie detection of this kind always shrinks the equilibrium set and if the probability for the truth or lie signal is sufficiently large the resulting equilibrium is unique. These results are robust to asymmetries regarding prior probabilities and payoffs.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

> "*Your poker face needs work my friend. It took me several seconds, but I can see now that you are lying.*" Dan Brown, The Da Vinci Code (2004, p. 553).

A recurring theme in fiction is that a character believes that others can see through his lies or that he can tell if someone is lying just by looking at or listening to him. In Dostoevsky's "Crime and Punishment" Raskolnikov is haunted by thoughts that police superintendent Nikodim Fomitj can tell that he is lying, and Sir Leigh Teabing in "The Da Vinci Code" believes, as the quote above indicates, that he has the ability to call a bluff. Furthermore, it is not uncommon that people in real life claim that they can see through a lie, while others claim that they are quite bad at bluffing.[1] In addition to this, psychological research suggests that although people in general are relatively bad at distinguishing lies from truths, they tend to believe that

☆ I have received valuable comments from Hans Carlsson, Toshiji Kawagoe and Tetsuo Yamamori and two anonymous referees. The original version of the paper was presented at the Economic Science Association's World Meeting in Amsterdam in June and at the international conference "Experiments in Economic Sciences" in Kyoto in December 2004. It has also been presented in seminars at the Department of Economics at Lund University and at the Department of Economics and Statistics at Göteborg University in Sweden. I am grateful for comments from the participants at these conferences and seminars. Financial support from Tore Browaldh and Jan Wallander foundation and The Bank of Sweden Tercentenary Foundation is gratefully acknowledged.

E-mail address: hakan.holm@nek.lu.se.

[1] Beliefs about such abilities might have had some historical impact. One example is when Hitler in a meeting before WWII lied to the ex-British Prime Minister Chamberlain about his intentions to invade Czechoslovakia. Chamberlain, obviously with a certain confidence in his abilities to identify liars, wrote to his sister: "in spite of the hardness and ruthlessness I thought I saw in his face, I got the impression that here was a man that could be relied upon when he had given his word" (Ekman, 1992, pp. 15–16). In the Parliament Chamberlain said that he was convinced that Hitler did not try to deceive him. Czechoslovakia was invaded by Germany a few weeks later.

0167-2681/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jebo.2010.06.003

Author's personal copy

certain observable non-verbal cues indicate lies (see e.g., Vrij, 2000, p. 58).[2] This psychological inclination motivates further investigations into the theoretical implications of lie detection cues in strategic situations. This paper will suggest a simple way to model truth and lie detection and demonstrate that detection cues may have important theoretical consequences in strictly competitive games.

Lying in strategic situations has received little attention in economic theory.[3] The standard assumption is that if a player does not want to, communication does not necessarily disclose his type or his intentions. Thus, lying is possible and costless. Without this assumption the literature on asymmetric information ought to be fundamentally modified. It can also be added that since disclosure of real intentions can be exploited by the counterpart, game theory (see Crawford and Sobel, 1982) predicts that zero-cost messages sent in games with conflict of interest are not informative.

Crawford (2003) and Hendricks and McAfee (2006) recognize that many strategic decisions (like where to attack in war) involve a stage where one party has the opportunity to misrepresent information. If players are heterogeneous with respect to their reasoning capacity, Crawford (2003) shows that misrepresentation may matter. However, there is also a more direct and psychological explanation that will be analyzed here. Players may actually be able, or believe that they are able, to recognize signals that are directly observed and related to the act of lying or truth-telling. Mathematically, this implies that the likelihood of detecting a lie is conditional on if the message is a lie or a truth, and furthermore, that the conditional likelihood of detecting a lie may differ from that of detecting a truth. The aim of this paper is to take a first small step in analyzing the implications of truth and lie detection.

Psychological research indicates that detection of truths and lies might differ. For instance, a review based on some 40 studies (see Vrij, 2000, p. 69) noted a 67 percent average accuracy rate for detecting truths. The corresponding accuracy rate for detecting lies was only 44 percent. Possible explanations are that truths are either easier to detect or believed to be associated with more frequent signals than lies.[4] Furthermore, lie and truth signals may be detected by new techniques. There are now several promising findings in neuroscience (see e.g., Spence et al., 2001; Langleben et al., 2002; Kozel et al., 2004) suggesting that certain simple lies (which involve inhibitory processes) are associated with higher activity in certain areas of the brain. The increased activity may differ between individuals, but is detectable by functional magnetic resonance imaging of the brain.[5]

In this paper it is not important if people in general are better at recognizing truths than lies or vice versa; the important thing being that beliefs about the recognizing capacity might differ and be conditional on whether the truth is told. One possibility is detection that is relationship specific. For instance, a man may know (or believe) that his wife sometimes with certainty can tell when he is lying and both are fully aware of this. The reason might simply be that the man is a hopelessly poor liar and that the wife has become a good lie detector after learning some observable cues associated with her husband's lies. When lying, the man is not in full control of when he emits these cues.

The issue of truth and lie detection is (to the author's knowledge) new in economics and game theory, and it would appear that analyzing truth and lie detection with game theoretical tools is something new in psychology. To investigate the effects of truth and lie detection, a simple signaling game is introduced and analyzed in Section 2. Implications of the results are discussed in Section 3 and the paper ends with some concluding remarks.

2. Theory

This section will analyze the implications of truth and lie detection. The class of signaling games, denoted as Bluffing games, can be presented as follows. First, nature selects the state $t \in \{B, W\}$. The probability for state B and W are p and $1 - p$, respectively. The S-player observes t and makes a statement $m \in \{B, W\}$ to the R-player about t. R is then to make a guess, $g \in \{T, F\}$, as to whether the message is true ($g = T$) or false ($g = F$). A statement is said to be true if $m = t$ and false otherwise. R wins, if her guess is correct, that is if $g = T$ and $m = t$, or $g = F$ and $m \neq t$. If R's guess is incorrect S wins. The players' payoffs are $u_S(t, m, g)$ and $u_R(t, m, g)$, respectively. Different assumptions about the payoff will be made, but for all variations it is assumed that $u_R(t, m, g) = 0$ when R loses and $u_S(t, m, g) = 0$ when S loses. Below, the impact of truth and lie detection will be analyzed in different sub-classes of Bluffing games.

[2] Even professional groups such as police officers who have to deal with lies on an everyday basis believe in stereotypical cues (like e.g., gaze aversion). It is shown by Mann et al. (2004) that such cues may mislead them in judging the veracity of suspects' stories.

[3] There is however, a growing number of interesting papers that experimentally investigate lying or issues closely related to it (see e.g., Frank et al., 1993; Dickhaut et al., 1995; Ockenfels and Selten, 2000; Blume et al., 2001; Brosig, 2002; Brandts and Charness, 2003; Charness and Dufwenberg, 2006; Gneezy, 2005; Cai and Wang, 2006; Hurkens and Kartik, 2009; Wang et al., 2010; Sánchez-Pagés and Vorsatz, 2007; Kawagoe and Takizawa, 2009). The approaches in these papers differ from the present one in that they involve reasoning about consequences, intentions, social preferences, or preferences for truth-telling.

[4] It should be mentioned that this difference can be explained in different ways. The latter explanation is consistent with the observation that people have a tendency to judge other's statements as truthful.

[5] Note, if lying creates physiologically different reactions from truth-telling, it is possible that this also generate differences in behavior that are (i) observable (ii) more easily recognizable for truths (or lies) than for lies (truths). For instance, Lubow and Fein (1996) found differences in pupil sizes under a "Guilty knowledge Test". Furthermore, voice stress analyzing software is also partly based on physiological different reactions. Despite that its reliability has been debated, this software has already been used by British insurance companies to screen telephone claims in the hope of detecting fraud (see New York Times, 2004, July 1st, "It's the Way You Say It, Truth Be Told", Technology section).

Author's personal copy

2.1. The Baseline case

In this case, S' and R:s payoffs as winners are $\bar{u}_S$ and $\bar{u}_R$, respectively. More, formally $u_S(B, B, F) = u_S(B, W, T) = u_S(W, B, T) = u_S(W, W, F) = \bar{u}_S$ and $u_R(B, B, T) = u_R(B, W, F) = u_R(W, B, F) = u_R(W, W, T) = \bar{u}_R$

Furthermore, it is assumed without loss of generality that $p > 1/2$.

We start by characterizing the set of perfect Bayesian equilibria. Let $m(t) \in [0, 1]$ denote the probability S that sends B after observing state t and $g(m) \in [0, 1]$ the corresponding probability that R guesses T after receiving m. Furthermore, $\mu(m) \in [0, 1]$ denotes the probability R believes that $m = t$ after receiving message m. This class of games is characterized by an infinite number of pooling and partly separating equilibria.

Observation 1. *The set of perfect Bayesian equilibria can be specified as follows*[6]:

(i) $m^*(\cdot) \in [0, 1]$, $g^*(B) = 1$, $g^*(W) = 0$ *and* $\mu^*(B) = p$, $\mu^*(W) = 1 - p$.
(ii) $m^*(B) \in (0, 1)$ *and* $m(W)^* \in (0, 1)$ *such that* $((1-p)m^*(W)/p) < m^*(B) < 2 + ((1-p)m^*(W) - 1/p)$, $g^*(B) = 1$, $g^*(W) = 0$, $\mu^*(B) = pm^*(B)/(pm^*(B) + (1-p)m^*(W))$ *and* $\mu^*(W) = (1-p)(1-m^*(W))/(p(1-m^*(B)) + (1-p)(1-m^*(W)))$.

Proof. When R plays $g^*(B) = 1$, $g^*(W) = 0$ S cannot increase his payoff since he will lose in the most probable state and win in the least probable one. Thus, all pooling strategies ($m^*(\cdot) \in [0, 1]$) are best-responses. When S plays a pooling strategy R:s best-response is $g(B) = 1$, $g(W) = 0$, which proves that (i) specifies equilibrium strategies. To understand that also (ii) specifies equilibria, note that even if S cannot improve on his payoffs by playing a separating strategy, there are limits on how much $m(B)$ and $m(W)$ can differ without inducing a different best-response from R. The upper and lower bound on the strategies is given by $(1-p)m^*(W)/p < m^*(B) < 2 + ((1-p)m^*(W) - 1)/p$. Furthermore, μ^* simply states the associated updated beliefs to these strategies.

We now prove that there are no equilibria in addition to ones stated. To do this, let us go through the possibilities other than $g^*(B) = 1$, $g^*(W) = 0$. Starting with pure strategies it is obvious that $g(\cdot) = 0$ or $g(\cdot) = 1$ cannot form an equilibrium since S can then always win by always telling the truth in the first case and by always lying in the second. Furthermore, the combination $g(B) = 0$ and $g(W) = 1$ would make S win in state B and lose in W independently of his strategy choice resulting in the expected payoffs $p\bar{u}_S$ and $(1-p)\bar{u}_R$, for S and R, respectively. However, this means that for all strategy choices of S, R can increase her payoff to $p\bar{u}_R$ by changing to $g(B) = 1$ and $g(W) = 0$.

Let us now consider the mixed strategies so that $g(B), g(W) \in (0, 1)$. Strategies such that $g(B) = g(W) \neq 1/2$ can be quickly be ruled out since the optimal response from S would be a pure separating strategy which is inconsistent with equilibrium. Furthermore, $g(B) = g(W) = 1/2$ cannot be an equilibrium since it will give R a payoff of $\bar{u}_R/2$. Thus, R can always improve on this and get $p\bar{u}_R$ by playing $g(B) = 1$ and $g(W) = 0$. When $g(B) \neq g(W)$ the only case for which not S would respond by a pure separating equilibrium is for $g(B) = 1 - g(W)$. However, with a mixed strategy combination it can be shown that for R:s pure strategy payoffs to balance in the information set where S has played B and W, respectively are $M(B) = (1-p)M(W)/p$ and $M(B) = (2p-1)/p + (1-p)M(W)/p$, respectively. This indicates an inconsistency. Q.E.D.

From Observation 1 it should be clear that R will play $g^*(B) = 1$ and $g^*(W) = 0$, since this means that R will always win (lose) in the most (least) probable state. As a consequence S can either play (i) an infinite number of pooling strategies $m^*(\cdot) \in [0, 1]$ or (ii) a more restricted set (but still an infinite number) of partly separating strategies.

2.1.1. Truth and lie detection

To simplify the presentation the detection will be phrased in terms of truths. However, lie detection can be modeled in a symmetrical way. Incorporating truth detection in this paper involves adding a stage in the game where R with a certain probability $\pi \in (0, 1)$ observes a perfect signal if the statement is true.[7] This stage takes place after S has made his statement and only if S tells the truth (i.e., $m = t$). The possibility of truth detection adds two (trivial) sub-games where R learns the true state. Let c denote the probability R plays T in these sub-games.

The introduction of truth detection here will refine the equilibrium, either partially or completely depending on how strong the truth detection is (in terms of π). Partial refinement means that some equilibria without truth detection will not be belong to the equilibrium set with it. We start by analyzing the weaker form of truth detection. The most important consequence is that S cannot play an infinite number of pooling strategies (i.e. the equilibria specified in *i* above) due to the truth detection.

Proposition 1. *If* $(2p-1)/p > \pi$, *the set of equilibria is given by*

(i) $m^*(\cdot) = 1$, $g^*(B) = 1$, $g^*(W) = 0$ *and* $\mu^*(B) = p(1-\pi)/(p(1-\pi) + 1 - p)$, $\mu^*(W) = (1-p)(1-\pi)/(p + (1-p)(1-\pi))$ *and the separating case*

[6] To simplify the presentation we leave out equilibria that differ only with respects to beliefs at information sets reached with zero probability. For instance, when S plays a pure strategy, a set of beliefs is consistent with equilibrium, for instance, $m(\cdot) = 1$, $g(B) = 1$, $g(W) = 0$, $\mu(B) = 1/2$, and $\mu(W) \in [0, 1/2]$.

[7] This signal could be a twinkle, a tick, a blush, a particular brainwave recognizable by EEG or a specific signal that the R player has learnt to associate with truth-telling. Furthermore, S knows this.

Author's personal copy

(ii) $m^*(B) > (1-p)/(1-\pi)p$, $m^*(W) = 1$, $g^*(B) = 1$, $g^*(W) = 0$ *and* $\mu^*(B) = p(1-\pi)m^*(B)/(p(1-\pi)m^*(B) + 1 - p)$ *and* $\mu^*(W) = 0$.

Proof.

Case (i) If $(2p-1)/p > \pi$ holds, $g(B) = 1$ and $g(W) = 0$ is still optimal for R and in state B S is indifferent between sending message B and W, since he will lose with both messages. However, in state W he will always win except for in the case he tells the truth and there is a truth signal. To optimize his expected payoff S must therefore play $m(W) = 1$. Hence, the only pooling strategy with a truth detection is $m(\cdot) = 1$.

Case (ii) Without truth detection the lower bound on S' truth-telling is given by $(1-p)m(W)/p < m(B)$. This lower bound changes to $(1-p)/(1-\pi)p < m(B)$ since $m^*(W) = 1$ and due to the truth detection signal. The upper bound was previously given by $2 + ((1-p)m(W) - 1)/p$, which is equal to 1 and always satisfied when $m^*(W) = 1$. The beliefs are updated with regard to the fact that $m^*(W) = 1$. Starting by a small truth detection $\pi = \varepsilon$ and then a somewhat larger, but still satisfying $(2p-1)/p > \pi$, additional equilibria can be ruled out by the same reasoning as in the proof of Observation 1. Q.E.D.

Note in case (i) truth detection works in the least probable state, W, in the intuitive way to deter S from truths. However, this will also force S to "tougher" restrictions on how little he can tell the truth in the more probable state (B) when playing a separating strategy. Hence, the effect of truth-detection on truth-telling might go in opposite directions in the different states. Nevertheless, it should be clear that even if truth detection is relatively weak in relation to the asymmetry in probabilities (i.e., $(2p-1)/p > \pi$) truth detection has a refinement quality in that pooling equilibria disappear and in the case of separating equilibrium, that S is forced to play $m^*(W) = 1$ and meets harder restrictions on truth-telling, since the lower bound on truth telling in B (i.e., $m^*(B)$) has increased from $(1-p)m(W)/p$ to $(1-p)/(1-\pi)p$.

We will now introduce truth detection that is strong in relation to the asymmetry in probabilities (i.e., $(2p-1)/p < \pi$). In this case truth detection leads to a unique equilibrium.

Proposition 2. *Introducing truth detection such that* $(2p-1)/p < \pi$ *leads to a unique perfect Bayesian equilibrium in the Bluffing game. The equilibrium is characterized by* $m^*(B) = (p - (1-p)(1-\pi))/p\pi(2-\pi)$, $m^*(W) = (p(1-\pi) - (1-p)(1-\pi)^2)/(1-p)\pi(2-\pi)$, $g^*(m) = (1-\pi)/(2-\pi)$, $c^* = 1$ *and* $\mu^*(\cdot) = 1/2$.

Proof. In this case pooling equilibria can be ruled out. Suppose, that $m(\cdot) > 0$, then R's updated belief concerning the probability of being in either node of the non-singleton information sets will differ. For instance, the probability of being in state B and W if $m = B$ is $p(1-\pi)/(p(1-\pi) + (1+p))$ and $(1+p)/(p(1-\pi) + (1+p))$, respectively. Due to the strength of truth detection, the latter probability is higher since $(2p-1)/p < \pi \Leftrightarrow 1 - p > p(1-\pi)$. Hence, it is more probable that S has lied about the state. It is then optimal for R to set $g(B) = 0$, which cannot be consistent with equilibrium, since S will react to it. In the case $m(\cdot) = 0$ a symmetrical reasoning can be applied on the non-singleton information set after $m = W$.

The weakly separating mixed equilibrium is derived by analyzing R's decision, where there are four different information sets, two non-singleton: IS_B and IS_W (after message B and W, respectively) and two corresponding singleton: IS'_B and IS'_W. At IS'_B and IS'_W, R has received a perfect signal, which means that it is optimal to set $c^* = 1$. At IS_B, R's conditional expected payoff from choosing T and F will be $\bar{u}_R m(B)p(1-\pi)/(m(B)p(1-\pi) + m(W)(1-p))$ and $\bar{u}_R m(W)(1-p)/(m(B)p(1-\pi) + m(W)(1-p))$, respectively.

These payoffs must be equal in a mixed equilibrium, which requires that $m^*(B) = m^*(W)(1-p)/p(1-\pi)$. A symmetrical reasoning can be applied for IS_W. Hence, R's expected payoff from choosing T and F will be $\bar{u}_R(1 - m(B))p/(1 - m(B))p + (1 - m(W)(1-p)(1-\pi))$ and $\bar{u}_R(1 - m(W))(1-p)(1-\pi)/(1 - m(B))p + (1 - m(W)(1-p)(1-\pi))$. By equating these two payoffs and by using $m^*(B) = m^*(W)(1-p)/p(1-\pi)$ in the substitution, we get $m^*(W) = (p(1-\pi) - (1-p)(1-\pi)^2)/(1-p)\pi(2-\pi)$. We can then also solve for $m^*(B)$ to get $m^*(B) = (p - (1-p)(1-\pi))/p\pi(2-\pi)$. S' strategy is adjusted to the fact that the (prior) probabilities of the states are different.

Now consider the balancing requirement on S' payoffs. Given $t = B$, S' payoff from the pure strategies of B and W are $\bar{u}_S(1-\pi)(1 - g(B))$ and $\bar{u}_S g(W)$, respectively and must also balance. Hence, $g^*(B) = (1-\pi)/(2-\pi)$. Consistent beliefs then require that $\mu^*(B) = 1/2$. A symmetrical reasoning can be applied for $t = W$, yielding $g^*(W) = (1-\pi)/(2-\pi)$ and $\mu^*(W) = 1/2$.

Uniqueness should be clear from the following observations. First, $\mu(m) \neq 1/2$ cannot be consistent with equilibrium since it would then be optimal for R to select a pure strategy in the non-singleton information sets, which leads to implications incompatible with equilibrium. For instance, $\mu(B) > 1/2$ must be combined with $g(B) = 1$. If this is the case R either plays this in combination with a strictly mixed strategy ($g(W) \in (0, 1)$) or a pure strategy ($g(W) = 0$). However, in the former case S would counteract by $m(B) = 0$, since this would give him a positive expected payoff compared to an expected zero payoff from $m(B) = 1$. This implies (due to the truth detection) that $\mu(W) < 1/2$ and hence that the optimal response is $g(W) = 0$ contradicting that $g(W) \in (0, 1)$. It remains to show that $g(B) = 1$ and $g(W) = 0$ is also inconsistent with equilibrium when $\mu(B) > 1/2$. When $t = B$, S is indifferent between the strategies. However, when $t = W$ optimality requires that $m(W) = 1$ due to the truth detection. But, this implies that $\mu(B) \leq 1/2$, which is inconsistent with the original assumption.

A pure separating strategy obviously cannot form an equilibrium in this "matching pennies" like game. Furthermore, if $\mu = 1/2$, then no strategy combination other than the one chosen will balance the expected payoff of the pure strategies. Together with the non-existence of any pooling equilibrium this establishes the uniqueness result. Q.E.D.

Author's personal copy

Clearly, the reasoning behind Propositions 1 and 2 also applies to lie detection, where it is assumed that π is the probability of R observing the perfect signal if the statement is false. Note, irrespectively of whether the detection is in terms of lies or truths, the introduction of it has a theoretical quality in that the equilibrium set is refined. Furthermore, the actual existence of detection signals is not crucial. What is crucial is that the players believe in them.

2.2. Robustness analysis

It is reasonable to ask how robust the results in Propositions 1 and 2 are to various changes. Consequently, the impact of truth detection in other classes of games than the one covered by the Baseline case is analyzed here. More specifically, the impact of truth detection will be analyzed when payoffs are state dependent and when there is simultaneous truth and lie detection. Each case is analyzed separately and the main intuitions behind the results are given.[8] The robustness analysis is conducted by introducing a single change at the time from the Baseline case, while all other assumptions are retained.

2.2.1. Case I: the sender has state dependent payoffs

In the Baseline case it is assumed that each player wins the same amount in both states. However, in some situations it is possible that it is more important for a player to win in one of the states. Therefore, the case where S' payoffs are state dependent is analyzed. Hence, it is assumed that: $\bar{u}_S^B = u_S(B, B, F) = u_S(B, W, T) \neq u_S(W, B, T) = u_S(W, W, F) = \bar{u}_S^W > 0$.

This modification in payoffs does not change the set of equilibria without truth detection compared to the Baseline case. The reason is that the payoffs for R (as a winner) are still the same in all outcomes, which means that R will continue to play $g^*(B) = 1$, $g^*(W) = 0$ to guarantee to win in the most probable state. S cannot change this and in each state he will be indifferent between his strategies, which means that all pooling strategies ($m^*(\cdot) \in [0, 1]$) belong to the equilibrium set. As before the asymmetry in prior probabilities will give room for a set of weakly separating strategies. The condition for these strategies will not be affected by S' state contingent payoffs since it is derived from optimality restrictions on R:s payoff. Truth detection does not alter the strategic situation either for the same reasons as above. Hence, truth detection with state dependent payoffs for S will generate the same set of equilibria as in the Baseline case, which means that Propositions 1 and 2 will hold without any modifications.

2.2.2. Case II: the receiver has state dependent payoffs

The analysis now turns to the case where R:s payoffs are state dependent. It is important to distinguish between the case where the payoff asymmetry goes in the same direction as the asymmetry in probability: $\bar{u}_R^B > \bar{u}_R^W$ and the case where it goes in the opposite direction $\bar{u}_R^B < \bar{u}_R^W$. In the first case the analysis and the equilibrium set without truth detection is basically the same as in the Baseline case. The only difference is that the payoff asymmetry reinforces R to play $g^*(B) = 1$ and $g^*(W) = 0$, which will make the conditions on the weakly separating strategies in (ii) somewhat less restrictive due to a $\bar{u}_R^W/\bar{u}_R^B$ term. Truth detection has a refinement effect in limiting S' strategy choices when it is weak and leads to a unique equilibrium when it is strong. The additional asymmetry does not affect this in any substantial way.

When $\bar{u}_R^B < \bar{u}_R^W$ the equilibrium set changes depending on if the asymmetry in probabilities dominates $p/(1-p) > \bar{u}_R^W/\bar{u}_R^B$ or if the payoff asymmetry dominates $p/(1-p) < \bar{u}_R^W/\bar{u}_R^B$. In the first case the equilibrium set without truth detection is exactly the same and the same as when $\bar{u}_R^B > \bar{u}_R^W$. Since the asymmetry in probabilities dominates, the best response to a pooling strategy for R is to win in the most probable state by playing $g^*(B) = 1$ and $g^*(W) = 0$. The equilibria with strong and weak truth detection will also be identical. However, when the payoff asymmetry dominates, R will exploit this by playing $g^*(B) = 0$, $g^*(W) = 1$ in equilibrium. This will lead to minor modifications of the equilibrium conditions that the strategies must satisfy. Similarly, some modifications are required to distinguish between weak and strong truth detection. However, truth detection will have the same qualitative refinement effects as in the previous cases.

2.2.3. Case III: simultaneous truth and lie detection

So far the analysis has only considered the possibility of either truth detection or lie detection. However, it is quite conceivable that in the same strategic situation some signals indicate truths and other lies. To take this into account the case with simultaneous truth and lie detection is analyzed. We therefore let $\sigma \in [0, 1]$ be the probability of R receiving a perfect lie signal when S is lying and without loss of generality analyze the case where truth detection is stronger than lie detection $\sigma < \pi$. The lie signal counter balances the truth signal, which affects the equilibrium specification slightly. However, the refinement results of Propositions 1 and 2 are qualitatively unaffected. The intuition for this result is that since $\sigma \neq \pi$ the presence of both truth and lie signals will not change the fact that not receiving a perfect signal will change the posterior probability of which node R believes she has reached in the non-singleton information sets.

Three cases have now been analyzed in addition to the Baseline case. The result of this analysis is summarized in Table 1. Qualitatively, the main refinement results are robust to the deviations introduced in all cases. Hence, truth detection has always at least partial refinement effects and, if sufficiently strong, it leads to a unique equilibrium.

---

[8] A more elaborate and formal analysis is provided in an appendix that is available online at: www.nek.lu.se/NEKJHO/BTAppendix.pdf.

Author's personal copy

Table 1
Characterization of how the results in the cases relate to Propositions 1 and 2.

Case	Payoffs	Both truth and lie signals	Partial refinement	Truth detection → unique equilibrium?
Baseline	Symmetric	No	Yes	If π is sufficiently large
I	Asymmetric (for *S*)	No	Yes	If π is sufficiently large
II	Asymmetric (for *R*)	No	Yes	If π is sufficiently large
III	Symmetric	Yes	Yes	If π is sufficiently large

3. Implications

One interesting implication of the result is that if someone could make *S* believe in the detection game, he would command a powerful tool to improve his information about the true state even if no perfect signals actually exist. Given that π is believed to be sufficiently large this holds for all cases analyzed in Section 2.2. To give a concrete example, suppose, before a large invasion by an *S*-army, a high-ranked officer in that army is captured by the *R*-army. Both armies know that *S* will soon attack, but only the *S*-officers know exactly where. There are only two feasible places to attack, Whitehall and Blackburn. *R* wins if and only if its officers succeed in deploying its army in the city attacked; the *S*-army wins otherwise. Now, assume that the captured *S*-officer only cares about the future success of his army and is brought in to be questioned by the *R*-officers about the place *S* plans to attack. Thus, the strategic situation is similar to the one in the bluffing game. However, before he is confronted with the crucial question he is successfully deceived into believing in a lie detection device possessed by the *R*-army and, also that the *R*-officers actually believe in it. He is informed that the probability that the lie detector will emit a perfect signal if he tells a lie is very high. When he is confronted with the crucial question Proposition 2 predicts that the officer will lie with a certain high probability. Hence, if the officer's answer to the crucial question is Whitehall, the *R*-officers will know that Blackburn is the probable target for the attack.[9]

4. Concluding remarks

The aim of the paper is to contribute to the study of truth and lie detection. It is noted that beliefs in signals revealing lies or truths are widespread, and that neuroscience has recently suggested new methods for detecting such signals. In this paper truth and lie detection is modeled as a probability that such a signal is perceived by the receiver. This probability is contingent on the truth value of the sender's message. The theoretical implication of lie and/or truth detection in bluffing appears to be an entirely new question in the literature and is given a first theoretical account here. The analysis is accomplished by developing a very simple symmetric bluffing signaling game with conflicts of interest. In such a game it is shown that the truth and lie detection can be incorporated into a game theoretical equilibrium analysis. The result of the analysis is that depending on the strength of the truth detection the equilibrium set shrinks either partially or fully to a unique non-pooling equilibrium. Given that the probability for the signal to be emitted is sufficiently large, this result is robust to various asymmetries in the players' payoffs. Furthermore, if someone is able to make the informed sender believe in truth or lie detection, this result could, in principle, be used to improve predictions about hidden information.

References

Blume, A., DeJoung, D., Kim, Y.-G., Sprinkle, G., 2001. Evolution of communication with partial common interest. Games and Economic Behavior 37, 79–120.
Brandts, J., Charness, G., 2003. Truth or consequences: an experiment. Management Science 49, 116–130.
Brosig, J., 2002. Identifying cooperative behavior: some experimental results in a prisoner's dilemma game. Journal of Economic Behavior and Organization 47, 275–290.
Cai, H., Wang, J., 2006. Overcommunication in strategic information transmission games. Games and Economic Behavior 56, 7–36.
Charness, G., Dufwenberg, M., 2006. Promises & partnership. Econometrica 74, 1579–1601.
Crawford, V., Sobel, J., 1982. Strategic information transmission. Econometrica 50, 1431–1451.
Crawford, V., 2003. Lying for strategic advantage: rational and boundedly rational misrepresentation of intentions. American Economic Review 93, 133–149.
Dickhaut, J., McCabe, K., Mukherji, A., 1995. An experimental study of strategic information transmission. Economic Theory 6, 389–403.
Ekman, P., 1992. Telling Lies: Clues to Deceit in the Market Place, Politics and Marriage. W.W. Norton, New York.
Frank, R., Gilovich, T., Regan, D., 1993. The evolution of one-shot cooperation: an experiment. Ethology & Sociobiology 14, 247–256.
Gneezy, U., 2005. The role of consequences. American Economic Review 95, 384–394.
Hendricks, K., McAfee, P., 2006. Feints. Journal of Economics & Management Strategy 15, 431–456.
Hurkens, S., Kartik, N., 2009. Would I lie to you? On social preferences and lying aversion. Experimental Economics 12, 180–192.
Kawagoe, T., Takizawa, H., 2009. Equilibrium refinement vs. level-k analysis: an experimental study of cheap talk with private information. Games and Economic Behavior 66, 238–255.
Kozel, A., Revell, L., Lorberbaum, J., Shastri, A., Elhai, J., Horner, M., Smith, A., Nahas, Z., Bohning, D., George, M., 2004. A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men. Journal of Neuropsychiatry and Clinical Neurosciences 16, 295–305.
Langleben, D., Schroeder, L., Maldjian, J., Gur, R., McDonald, S., Radland, J., O'Brien, C., Childress, A., 2002. Brain activity during simulated deception: an event-related functional magnetic resonance study. Neuroimage 15, 727–732.

[9] It should be mentioned that it is important that the players are sufficiently rational, but not too smart. For instance, if the captured officer realizes the deception scheme and can feint that he believes in it, he can use a meta-feinting strategy by telling the truth since he knows it will be taken as a lie. Obviously, there is no end to such a meta reasoning.

Author's personal copy

Lubow, R., Fein, O., 1996. Pupillary size in response to a visual guilty knowledge test: New technique for the detection of deception. Journal of Experimental Psychology: Applied 2, 164–177.
Mann, S., Vrij, A., Bull, R., 2004. Detecting true lies: police officers' ability to detect suspects' lies. Journal of Applied Psychology 89, 137–149.
Ockenfels, A., Selten, R., 2000. An experiment on the hypothesis of involuntary truth-signalling in bargaining. Games and Economic Behavior 33, 90–116.
Sánchez-Pagés, S., Vorsatz, M., 2007. An experimental study of truth-telling in a sender-receiver game. Games and Economic Behavior 61, 86–112.
Spence, S., Farrow, T., Herford, A., 2001. Behavioral and functional anatomical correlates of deceptions in humans. Neuro Report 12, 2849–2853.
Vrij, A., 2000. Detecting Lies and Deceit, The Psychology of Lying and the Implications for the Professional Practice. John Wiley& Sons, Chichester.
Wang, J., Spezio, M., Camerer, C., 2010. Pinocchio's pupil: using eyetracking and pupil dilation to understand truth-telling and deception in games. American Economic Review 100, 984–1007.