HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

Deterrence Games and the Disruption of Information

Faculty Research Working Paper Series

Siyu Ma
Tel Aviv University

Yair Tauman
Stony Brook University

Richard Zeckhauser
Harvard Kennedy School

August 2020
RWP20-026

Deterrence Games and the Disruption of Information

Siyu Ma[1 *] Yair Tauman[2 †] Richard Zeckhauser[3 ‡]

August 5, 2020

Abstract

Deterrence is a generic situation, where player 2 threatens to bash player 1, should he harm her. Typical player 1's are bomb builders, market invaders, or computer hackers. 2's decision whether to bash will depend on a noisy signal. 1 may have the ability to disrupt the signal. The game has a unique equilibrium. If it involves mixed strategies, signal reliability benefits both players. Signal disruption hurts 2 for sure and only benefits 1 if 2 believes he is unlikely to possess it.

Keywords Deterrence, Intelligence, Information Disruption

1 Introduction

Powerful players, be they nations, companies, or individuals, must always guard against a lesser player who could take actions that harm them. For example, an enemy nation may wish to join the nuclear club; a hostile country may want to hack their technology; a startup company may be seeking to grab market share from a market-dominant firm by innovating around its prime patent.

In some situations, the powerful player, Player 2, can prevent such actions. In others, those that concern us here, her principle defense is deterrence. She will threaten to severely hurt the

[*1] School of Mathematical Sciences, Tel Aviv University
[†2] Adelson School of Entrepreneurship, Interdisciplinary Center (IDC), Herzliya and Economics Department, Stony Brook University
[‡3] Harvard Kennedy School, Harvard University.

lesser player (Player 1) should he take the harmful action, for example by attacking his bomb-building facilities. More generally, a deterrence game arises whenever Player 1 can take an action that hurts Player 2, but Player 2 has the ability to retaliate decisively against Player 1. Hereafter, for brevity, we will use 1 and 2 to indicate Player 1 (male) and Player 2 (female).

Table 1 discusses some representative scenarios. The prime example we employ is of the small company, 1, seeking to grab part of the market of its major competitor by innovating around its patent. Its major competitor, 2, could crush such an effort by launching a substantial R&D effort in the same realm, but it would not want to do so if the small company was not seeking a design-around. The major company will monitor its small competitor, and the small competitor will seek to evade such monitoring, for example by not presenting early results at conferences or by strengthening its computer security. In most cases discussed below it will seem more intuitive to regard 2 as more powerful than 1, given her strong deterrent capability. Hence, we will usually posit that relationship to facilitate exposition. However, the analysis is perfectly general, and 1 could be equally or more powerful than 2. For example, the firm seeking to grab market share could be a technological behemoth whose harmful action might also include innovate-around patent efforts or, alternatively, buying and investing heavily in a rival to the market-dominant firm.

In a wide variety of circumstances, 1's action can be taken in secret. If 2 does not discover the harmful action reasonably swiftly, her retaliatory action will come too late. The bombs will have been built and effectively hidden; too much infrastructure, e.g., computers or electrical grid, will have been undermined; the startup's market position will have been established.

To avoid reacting too late, 2 needs an intelligence system. That system, which is usually imperfect, will enable her to bash on a timely basis; that timely capability is what enables her to deter. Beyond timeliness, there is a second concern about effective deterrence. 2 wants to retaliate if and only if 1 has taken the harmful action. Misguided retaliation is costly to both parties. Of course, the imperfection of the intelligence weakens the ability of 2 to retaliate appropriately, enabling errors of both Type I (not retaliating when 1 has taken the harmful action) and Type II (retaliating when 1 has not taken the harmful action). 1, of course, likes best the outcome where he takes the harmful action but escapes retaliation. To make this outcome possible, not only does he choose in secret, but he also can employ a technology to disrupt 2's intelligence system. The prime focus of this paper is on the consequences of this disruption capability. Its surprising general finding is that 1's ability to disrupt is often detrimental to the interests of both player 1 and player 2. That misfortune is compounded because player 1 will rarely have a credible ability to foreswear his use of his disruptive technology. We should be clear at the outset that the disruptive technology need not be some high-tech operation. For 1 as a potential patent innovator, it could be as simple as not participating in conferences where technological advances are reported.

Though deterrence games are discussed most commonly in the context of international affairs, they arise regularly across a broad swathe of situations ranging from business to romance.

Most commonly, 1 is less powerful than 2, but in the jiu jitsu of life, the weaker player often enjoys compensatory advantages. And there are situations of mutual deterrence between two players, each of whom could severely retaliate against the other, despite power asymmetries. This was the case, for example, with the United States and the USSR in the later stages of the Cold War.

We study deterrence situations in abstract form, presenting examples mainly at the outset. 1 has the ability to take an action, H, that would *harm* 2. 2 has an action available that would *bash*, B, 1. That is, 2 has a threat capability against 1 (Schelling, 1960, pp. 35-43). The two players' alternative actions, respectively, are *not harm*, NH, and *not bash*, NB.

In this analysis, threats are assumed to be credible, whether to maintain 2's reputation, or simply due to pure preference quite apart from reputation[1]. As an example of pure preference, the United States would probably prefer to destroy North Korea's nuclear bombs and facilities.

Of course, the threatening party, 2, would far prefer that the harmful action not be taken, namely NH. Moreover, if it is not taken, NH, it would far prefer not to bash, NB, inappropriately. That is, $(NH, NB) \succ (NH, B)$. In addition, if the harmful action is taken, bash is the favored action. That is $(H, B) \succ (H, NB)$. This inequality makes the threat credible. Finally, justified bash is superior to an unjustified bash, namely $(H, B) \succ (NH, B)$. That could be because the powerful entity is greatly concerned with its legitimacy. Such legitimacy would be damaged should it be discovered later that she punished without justification.

The preferences of 1, the potential harmer, are $(H, NB) \succ (NH, NB) \succ (NH, B) \succ (H, B)$. Suffering a bash is always costly, but being bashed when there is justification is worse. His preference $(NH, B) \succ (H, B)$ arises in part because H, say building a nuclear bomb or designing around a patent, is expensive. Moreover, quite apart from cost, H would do 1 little good if he gets bashed. Finally, he would welcome 2's loss of reputation should she bash without justification. Note, 1's preferences given bashing are the reverse of those for the powerful player.

Uncertainties are everywhere in this game, so probabilistic assessments based on Bayesian updating play a critical role. Thus, each player employs von Neumann-Morgenstern utilities, and their values are common knowledge.

We tend to use the patent and market-grabbing example in this analysis, but either of the other two situations, or a dozen others that the reader might provide, would lead to the same

[1] In a major variant of Schelling's threat game, the powerful player, whether a parent threatening no dessert should a child not complete her homework or a nation threatening nuclear attack should an enemy launch an invasion, would prefer not to carry out her threat. In such a situation, the threatening party, 2, makes a public commitment to bash 1 should he take action H. Such a public announcement puts her reputation at risk and lowers her payoff significantly should she fail to bash in response to H (Schelling page reference). Thus, she prefers B to NB given H. That is, her public commitment makes her threat credible.

Table 1: Scenarios with Harmful Actions for Player 1 and Bashing Actions for Player 2

Actors	Harmful Action	Bashing Action
Companies	Grab market share (Design around patent)	Preempt theft (Improvement patent)
Nations	Build nuclear bomb	Attack bomb facilities
Nations/Companies	Hack computers	Massively hack computers

analysis. The table only identifies one class of bashing actions. Our literature review discusses several others, and the real world reveals a vast array.

Note, 1 would seek to take a harmful action on a clandestine basis, knowing that 2 would not punish if she thought that NH prevailed.

2, who understands that secret preparations are highly desirable for 1, will develop an intelligence system to monitor 1's actions. For example, the United States continually tries to assess weapons developments in Iran and North Korea.

Patent-based companies constantly try to determine whether former employees might be innovating around their patents. And companies and nations vigorously and continuously scan the Internet as a critical component of their efforts to detect the theft of intellectual property. However, all such detection efforts are far from perfect. The result is a cat-and-mouse situation. 1, knowing that he is being monitored, seeks to hide any harmful actions more effectively. He casts a web of silence around any research efforts on a patent work around. He builds bombs underground. He disguises his hacking efforts as coming from third countries. He writes his plans for developing harmful technologies in code. He builds sturdy firewalls around his computers. Such strategies disrupt the capability of 2's intelligence apparatus.

This analysis studies how the capabilities of 2's intelligence and 1's disruptive technology affect the two players' actions and their expected payoffs. The finding is important, because the deterrence game is played every day in dozens of contexts throughout the world. Moreover, 2 and 1, operating in whatever realm, make significant efforts to deploy and disrupt intelligence.

The game proceeds as follows: Initially, 1 chooses his action, H or NH, in secret. He simultaneously decides whether to disrupt the major company's intelligence capability. 2 receives a signal on 1's action. That signal may have been disrupted, but 2 would not know if it had. Then 2 chooses her action, either following that signal, going opposite to it, or ignoring it (choosing either B or NB regardless of the signal). Payoffs, depending on the two players' actions, are then received, and the game is over. The signal, whether or not disrupted, does not affect payoffs directly. Indirectly, , of course, it can affect 2's action. The structure of the game, the players' payoffs, and the performance of the intelligence system with and without disruption, are common knowledge.

The only asymmetry of information is about whether 1 has a disruptive technology. 1 knows the quality of his disruptive technology, but 2 only has an unbiased prior on its value. That is, there are two types for Player 1, 1^+ and 1^-, where 1^+ possesses the technology and 1^- does not. 2 knows the probability, g, that he faces 1^+ but not whether he actually faces 1^+ or 1^- (in case $0 < g < 1$).

To illustrate, let us elaborate on the grab market share situation. 1 is a "small" start-up company headed by a former employee of 2, a "major" company with a highly successful patent-protected market position. 1 would like to grab a portion of 2's market by innovating around its major patent (the action denoted by H). 2 can thwart this effort by investing heavily in the technology for an improvement patent; that would represent action B. Given 2's far greater R&D capability, she would almost surely be successful and 1's efforts to grab with an innovate-around would have been wasted. However, if 1 had been deterred from the grab strategy (taken the action denoted by NH), 2 would have preferred to have chosen NB. Pushing R&D for an improvement patent would have been mostly a waste of money.

2 has an imperfect intelligence system that sends one of two signals: h to indicate that 1 is trying to grab market share by innovating around her patent, and nh to indicate that 1 is not seeking to grab. 2's intelligence system would include monitoring hires, talking to industry sources, scrutinizing conference attendance, and hiring an investigative company whose methods are secret. The precision of 2's intelligence is α, the probability that it informs correctly about 1's action. 1 has a disruptive technology. For example, it might add a layer of misinformation as a means to protect its information technology system.

The disruptive technology renders 2's intelligence worthless with probability β. Thus, β represents the disruptive technology's quality. The disrupted intelligence system sends the signals h and nh each with probability $\frac{1}{2}$, independent of the action taken by 1. Based on the signal and on her belief on the action taken by 1, Player 2 decides whether to bash 1, B, or not bash, NB.

The initial framework of this game is similar to Biran and Tauman (2009) and Jelnov, Tauman and Zeckhauser (2017) (BT and JTZ1 hereafter), which both dealt solely with the bomb-building scenario. The preferences of the players here are the same as in BT. The major innovation here is to allow for 1 to possess a disruptive technology.

This paper seeks to add to the general literature on deterrence theory. Hence, it does not focus on the bomb-building case. It offers a conceptual advance over BT and JTZ1 by incorporating the possibility that 1 can disrupt the signal that 2 receives from her intelligence. In a concluding section, it distinguishes between preemptive deterrence, as for example in classic entry deterrence models in industrial organization, and after-the-fact deterrence, as in most military threat situations. It also shows that both of these should be thought of as a three-step game, not the two-step game that is normally analyzed. Thus, it adds an intelligence and decision making step at the end of preemptive deterrence models, and places a step to create an intelligence and/or bashing capability at front of after-the-fact deterrence models.

Our game has a unique equilibrium. Depending on the (seven) parameters that define it, it can be either a mixed strategy equilibrium or a pure strategy equilibrium. In a mixed strategy equilibrium, 2 employs a mixture of just two pure strategies. One of them follows the signal. The other is either to bash or not bash, regardless of the signal. 1 has two choices to make: should he take a harmful action and should he use his disruptive technology in case he has it. Hence, he has four possible strategies. At the equilibrium he mixes two of those strategies: harm and use the disruptive technology, or not harm and not use the technology. The other two combinations are never used. In a pure strategy equilibrium, 2 just follows the signal. She bashes if the signal is h and does not bash if the signal is nh. If 1 has a disruptive technology, he will for sure harm 2 and if he does not have the technology he will not harm her.

Not surprisingly, regardless of the parameters Player 2's payoff is increasing in the precision (α) of her intelligence. More surprising is that 1 also benefits from a greater precision of 2's intelligence. That is because lowering the intelligence's efficiency triggers 2 to act more aggressively by bashing 1 with higher probability, which hurts 1. A similar result is already derived in BT and JTZ1.

Suppose 1 possesses a disruptive technology. How does this technology affect both players' payoffs? Intuition might suggest that the better this quality (a higher β), the lower the payoff to 2 and the higher the payoff to 1. While 2 does suffer from 1's better disruptive technology (which is consistent with the previous paragraph), 1 may suffer as well if 2 believes that he possesses the technology with high probability.

The reason is that as 2 strongly believes that her signal will be disrupted, she can discern fewer of 1's actions and therefore expect 1 to take advantage by harming her more readily. In response, 2 bashes more readily, lowering the payoff of 1. In short, a better disruptive technology induces both players to be more aggressive. This result suggests that 1 would be better off scrapping his technology if he can credibly convince 2 he has done this. On the other hand, if 2 believes that 1 possesses the technology with relatively low probability, both players will be less hostile. Then 1 can benefit from his technology. In other word, possessing a disruptive technology is only beneficial for 1 if he can hide it quite well.

As mentioned above, Player 2, irrespective of the parameters of the game, always plays "follow the signal" strategy with positive probability (possibly always 1). This, together with the imperfect signal, implies that she bashes 1 with positive probability whether 1 is innocent or not. If one takes seriously the equilibrium concept as a mode of behavior, then in such conflicts 1 who has disruptive technology will harm 2 with positive probability (possibly 1). 1 who does not have this technology will play NH with positive probability (possibly 1). Since 2 always follows the signal with positive probability, it is also possible that 1 goes unpunished even when he has harmed.

An important implication of this paper is that developing a disruptive technology is beneficial for 1 if he makes some pre-play efforts to deceive 2, lowering 2's belief of him having it. In this case, a better quality of 1's technology increases his payoff. If, however, 2 strongly and correctly believes 1 has a disruptive technology, he is worse off having it. Moreover, the better quality of his technology decreases his payoff.

2 Related Literature

By far the most extensive discussion of deterrence, and the challenges created by asymmetric information, has been in relation to military affairs, and that is where we start. Deterrence of military aggression has been with us since ancient times (Cioffi-Revilla 1999). However, as technologies have changed, so too has the nature of deterrence. In olden days, a nation having sufficient military capability could deter an enemy from invading by threatening to deliver punishing blows should he do so. Given the development of weapons of massive destruction, deterrence became more of a tit for tat measure: You hit me, and I will hit you hard enough to render costly your initial hit.

The stability created by massive weaponry on both sides was undermined by the possibility that a first strike would wipe out most of the other party's weapons. Such concerns gave rise to hiding and moving weapons and creating essentially invulnerable nuclear-armed submarines. Those steps in turn led to the realization that making A's weapons more secure actually made his enemy B safer, because it would reduce the incentives for a first strike. That initially unintuitive result, like those preceding it and a number in this paper, made it important to study deterrence situations from a game-theoretic perspective.

Surprising as it may seem in retrospect, it took many brilliant mind years to understand the underpinnings of the deterrent effects of massive retaliation. As Schelling's The Strategy of Conflict (1960, p. 7) observes: "What is impressive is not how complicated the idea of deterrence has become and how carefully it has been refined and developed, but how slow the progress has been, how vague the concepts still are, and how inelegant the current theory of deterrence is." Schelling, whose work, deeply informed by strategic thinking, ultimately won the Nobel Prize, developed his own analyses using minimal mathematics.

An early extensive treatment of nuclear deterrence using mathematical formulations of game theory is due to Powell (1990). He explores credibility deeply, as does Schelling (1960). Fearon (1995), whose celebrated article "Rationalist Explanations for War" eschews mathematics though his appendix employs it, addressed the related question of why wars, which obviously entail deadweight losses, occur. He identifies the potential of going to war as a rational action.

Asymmetries of information, a centerpiece of this analysis, provide one of his two primary explanations. Lack of commitment capability is the other. Schelling (1960) also devotes extensive attention to commitment capabilities and asymmetric information. His later writings placed such findings firmly in policy contexts. Schelling strongly supported the sharing of information between the United States and the USSR as a means to avoid war.

Recently, a new generation has shown how game theory, and its sidekick Bayesian analysis, combined with deep thinking, can yield important insights into both effective deterrence and mistaken attacks. Debs and Monteiro (2014, p. 1) address secret build ups, preventive wars, including mistaken preventive wars: "states may be tempted to introduce power shifts as a fait accompli. Knowing this, their adversaries may strike preventively even without conclusive evidence of militarization. In fact, the more effective preventive wars are, the more likely they will be launched against states that are not militarizing." (The fait accompli potential of the harmer is also a feature of entry deterrence models.) They employ the Iraq War to illustrate. It stemmed, they observe, from Iraq's inability to commit not to develop nuclear weapons combined with the United States' inability not to launch a preventive war. Bas and Coe (2012) further develop the model of a potential power shift as a trigger to war. They illustrate with historical examples from precolonial New Zealand to the 1967 Six-Days War.

Duelfer led the Iraq Survey Group's considerably more massive effort to assess a mistaken preventive war. He subsequently attributed the Iraq War, disastrous to both sides, to chronic misperceptions (Duelfer and Dyson 2011, p. 80). The United States and Iraq even misperceived which nation Iraq thought to be its enemy. Saddam Hussein thought Iran; the United States thought the United States and allied nations. Hussein deceived foreign intelligence and indeed his own army into believing that he possessed WMDs. He felt "the need to maintain the outward appearance of possessing a WMD capability to deter Iran." (Duelfer and Dyson 2011, p. 94). Finally, Hussein thought the threat by the US-led coalition to go to war unless the UN resolutions were met was a bluff. Consistent with the current paper, each side may have behaved rationally, given its beliefs.

Information is at the heart of most analyses of deterrence, and related military subjects, such as arms control. Coe and Vaynman (2020, p. 342) highlight the importance of information in any attempt to limit armaments: "The main impediment to arms control is the need for monitoring that renders a state's arming transparent enough to assure its compliance but not so much as to threaten its security." Effective monitoring is thus critical, lest cheating be worthwhile. Credibility and deterrence show once again their links. Here it is through information, whereas for Schelling and Powell it was a feature of payoff structure.

Signal disruption, a key feature of this analysis, impedes an opponent's ability to detect one's type; it is a form of deception. Deception gets featured in a number of contemporary studies. For example, Slantchev (2010) considers a model where an initial pre-crisis bargaining stage precedes a possible war. A bargained outcome will favor a strong player. Hence, a strong player may signal strength by making a high demand for a negotiated outcome. But she mixes strategies and sometimes feigns weakness by making a modest demand to lure her opponent into failing to arm.

Refusing arms inspections is another way to deny information to a potential basher. Baliga and Sjostrom (2008) show that the strategic ambiguity refusing introduces can deter a basher from attacking. Thereby, it can serve the same role as possessing WMDs. They also show, unfortunately, that ambiguity can lead to mistaken attacks. Finally, they provide an insightful analysis of the possible roles for direct communication between the players.

Our model relates broadly, though more distantly, to a great variety of models in the arms building, nuclear deterrence, and arms control fields, and more generally to military strategy. O'Neill (1994) provides an extensive survey of this literature.

The underpinnings of our model are drawn from BT and JTZ1. Both address deterrence situations where intelligence plays a salient role. Those prior contributions have Player 2 act cautiously if her intelligence is relatively accurate, namely above a cutoff defined by the parameters of the game. She never bashes if she receives the signal nh. If she receives the signal h, she mixes the strategies bash and not bash. If accuracy is below the cutoff, she plays aggressively, always bashing given signal h, and mixing the two strategies given the signal nh. This paper places those results in a larger and more realistic context. It evaluates the players' strategies and outcomes when Player 1 may possess a disruptive technology. Such technologies are a common weapon in real world contexts where the other player is deploying intelligence. As expected, Player 2's choice whether to act cautiously or aggressively depends not only on the intact precision of intelligence, but also on Player 1's ability to disrupt the intelligence. A surprising result is that 1's disrupting ability actually lowers his payoff if his ability is commonly known. The reason is that it shifts the strategy of 2 and makes her more aggressive. Hence, if 1 can do so credibly, he should commit not to using his disruptive technology. This is true if 2 knows for sure that Player 1 possesses disruptive technology and the result is reversed if 2 believes that 1 is likely not to have disruptive technology.

Formally we consider the case where there are two types for 1. One has the ability to disrupt; the other does not. Player 2 only knows the probability that she faces either type. In this asymmetric information case, 1 loses from having the disruptive technology (of any quality) if 2 believes the probability of being the disruptive type is above some cutoff. 1 benefits if 2 believes the probability is below that cutoff.

Jelnov, Tauman and Zeckhauser (2018) (JTZ2 hereafter) also consider two types for Player 1. However, its types differ on preferences, not capabilities. There additional type for 1 is a provocateur. Unlike players 1 in this analysis or JTZ1, a provocateur prefers to be bashed when he has played NH, due to the support he would receive and the blame 2 would get from an unjustified attack. Provocateurs are not part of this analysis. Here, 1 always prefers to avoid an attack.

Patents represent a second area where deterrence and bashing are brought into play. Patents can be used strategically by the patent holder as entry barriers, especially in industries like biotechnology or software where patent registration and renewal occur frequently. Firms make their patenting decision strategically for different purposes: to maintain or establish their position in a technological domain, to block rivals from patenting related inventions, to expand their portfolio even with lower quality patents as a defensive strategy (Hall and Ham-Ziedonis, 2001) or to improve bargaining positions in negotiations with other firms (Cohen, Levin and Walsh, 2000).

In all these situations, a patent holder must worry about a potential entrant or patent-weakened competitor innovating around its patent. The theoretical literature on using patents to deter entrants dates back to the 80s. An array of sophisticated deterrence strategies have been studied. Gallini (1984) shows that an incumbent firm may strategically offer a fixed fee license to a potential entrant to reduce its incentive to develop its own, possibly better, technology. The incumbent firm, threatened by the risk of successful research by a market challenger, can therefore secure a market share by giving up its monopoly power, in order to remain the technological leader. Rockett (1990) examines licensing as a means of choosing between a weak and a strong competitor, which a patent holder, the monopolist, will face once the patent expires. The patent holder prolongs its dominant position by licensing the weak competitor before expiration, and thereby keeping out the strong competitor even after expiration. Subsequent work has built on the approach of helping weak competitors to deter strong competitors. Ellison and Ellison (2011) (hereafter E&E) discuss the deterrence behavior of monopolist pharmaceutical firms in the period just prior to drug patent or FDA exclusivity expiration. Using data from 63 cases of chemical compound patents facing imminent expiration, they find examples of incumbent firms decreasing advertising slightly prior to patent expiration. In intermediate-sized markets where generic drug entry is imminent, the incumbent firm may prefer to retain its status as a monopolist by allowing the market to shrink, rather than submit to competition in a more robust market. In our terminology, they gain by pre-bashing the profits of a potential entrant.

In securing the ability to bash, superior knowledge is a powerful tool. We would expect a patentholder to be better informed about demand than her competitor. Milgrom and Roberts (1982) show how a monopolist may use a “limit pricing” strategy to deter market entry. This strategy of pricing a product below its profit-maximizing level to deter potential entrants with lesser information to believe the market is less profitable than it is. These information-deprived potential entrants may interpret the low prices as a signal that the incumbent has low costs, hence a major advantage. This strategy will fail, as Harrington (1986) shows, if potential entrants don’t know their costs, but believe that they are strongly correlated with those of the incumbent. In their investigation of pharmaceutical firm behavior, E&E find evidence of price signal disruption. Noting that prices and revenues tend to be industry-wide common knowledge, they theorize that signal disruption is used to confuse potential entrants about consumer elasticity of demand.

Entry deterrence by patent holders is a subset of such deterrence by monopolists. Klemperer (1987) shows how an incumbent firm can protect itself against competitor market entry by building brand loyalty. This strategy can be effective even when the potential competitor has product offerings of similar quality, so long as consumers perceive "switching costs" to be high. E&E found evidence of loyalty-building efforts in their pharmaceutical data. Shortly before patent expiration, they found that pharmaceutical firms tended to expand their dosage offerings and drug delivery methods. Due to laws regarding pharmacological dispensation, firms wishing to offer generic drugs must match the dosing options of the incumbent to remain competitive. This raises up-front production costs. For example, consumers who prefer a generic drug may be denied filling a prescription with two generic 50 mg capsules rather than the one brand-name 100 mg capsule that was prescribed. Calling the prescriber to approve the dosing in favor of the generic alternative is often a sufficient deterrent to adopting the generic option.

Sophisticated pricing strategies can also be used as a bashing weapon when firms do enter. If consumers value authenticity, firms can bash counterfeiters by raising their prices. Qian (2008) measured a 45% average price increase within two years of infringement in Chinese markets.

Firms subject to counterfeiting's closely related cousin, piracy, have directly retaliated against infringers. A famous bashing effort was undertaken by the heavy metal band Metallica. In 2000, it sued peer-to-peer music sharing service Napster for copyright infringement. Shortly after the lawsuit was filed, Napster agreed to ban the 300,000 users from its platform who had illegally downloaded Metallica content (National Post 2000). Shortly after this move, smaller firms. lacking the capital to file lawsuits, found other ways to bash the music platform and its users for piracy. They uploaded "cuckoos eggs" – corrupted song files – to the website, and thereby eroded Napster's credibility amongst users (New York Times 2000). Larger music producers even sought legislation granting exemptions to antihacking laws, allowing them to directly attack copyright-infringing peer-to-peer networks. Metallica's lawsuit is cited as igniting a broader pushback that led to the demise of once mighty Napster (Wall Street Journal 2002). Such an example illustrates that the strategy of deterring threatened harm with threatened bash may occur in lesser-regulated intellectual property environments. Moreover, many lighter bashes by smaller entities may have a collective cumulative effect to deter or remedy harmful actions, such as counterfeiting or piracy.

Barrachina, Tauman and Urbano (2014) analyzes business espionage (intelligence) in the context of industrial entry deterrence. They consider a monopoly incumbent, who may expand capacity to deter entry, and a potential entrant who owns an imperfect intelligence. The intelligence generates a noisy signal on incumbent's actions and the potential entrant decides whether to enter based on this signal. If the precision of the intelligence is commonly known, the incumbent will benefit from his rival's espionage. In contrast, the spying firm (the entrant) will typically gain if the espionage accuracy is sufficiently high and privately known by her. The result of symmetric information case is consistent with our model, but we further consider and emphasize the impact of signal disruption on equilibrium.

Cyberattacks are a third policy realm where deterrence and bashing come into play. Cyberattacks represent a low-cost way to impose significant harm on another player by gathering intelligence, stealing intellectual property or impairing infrastructure. Large-scale attacks and operations are usually undertaken by state governments, though smaller attacks may be perpetrated by criminal groups seeking ransom, companies stealing information, or terrorist groups just wanting to harm. Our focus is on attacks supported by nations hoping to harm their enemies. Compared to conventional warfare, cyberattacks to date have imposed little physical damage and virtually no fatalities, but this may change.

An early consequential attack on physical infrastructure was the Stuxnet cyber worm, which effectively targeted Iranian nuclear facilities in 2009 and 2010 (Chen 2010; Stevens 2020). This suspected joint effort by the United States and Israel was a preemptive strike that slowed Iran's nuclear program (Sanger 2012). In 2015, suspected Russian hackers triggered a power outage in Ukraine affecting 225,000 customers (Lee, Assante, and Conway 2016). The Ukraine's power grids were targeted again a year later (Slowik 2018). In 2017, a Saudi Aramco petrochemical plant was attacked in a manner likely intended to cause an explosion. The culprit(s) have not been identified, though government support is assumed (Perlroth and Krauss 2018). Cyberattacks on critical industrial control systems have become an increasingly accessible tool for nations to use against each other. They blur the lines between cyber- and conventional warfare.

Governments and companies spend large amounts to protect their computer systems, but attackers continually find new ways to invade and disrupt them. Hence deterrence through the threat of bashing has been brought into play. A salient example, consistent with the themes in this paper, occurred in spring 2020. Iran broadly hacked into the computers of Israeli water systems; little damage was done (Warrick 2020a). Israel bashed in return by launching a cyberattack on a major Iranian port; shipping was delayed for days (Warrick 2020b). Bigger attacks of the same kind are unlikely to be useful against cyber theft. In the Huawei case, the United States retaliated against specific individuals from the company it accused of theft, and made major efforts to hinder the company from securing technology and selling abroad.

Arms control agreements, the threat of devastating retaliation, and the fact that attackers are easily identified have prevented wars between major powers over many decades. The combination of international agreements and deterrence has worked. No so for cyberattacks. A decade ago, when cyberattacks were in their infancy, there was already widespread agreement that cyber arms treaties were unlikely to develop, due primarily to problems with attribution and widespread accessibility of cyber weapons (Dipert 2010, 406). To date, there has been minimal international regulation. UN Secretary General Antonio Guterres noted that "cyber warfare is already happening – in a lawless international environment" ("Secretary-General's Video Message to Forum of Small States" 2020). Indeed, the UN has a longstanding multi-country group within the Office for Disarmament Affairs (UNGGE) assigned to address cyber warfare; in its most recent 2017 iteration, it was unable to come to a consensus and release a report as it had done in prior years. International consensus on cyberwarfare may take years to come (Henriksen 2019)[2].

Absent effective international agreements, which also help to establish taboos against the use of particular weapons (Schelling 2005), deterrence will have to serve as a prime defense against cyberwarfare. Existing theories of deterrence were developed in the strategic context of the Cold War. The goal of nuclear weapons is to possess a capability that accomplishes its purpose without ever being used. Deterrence with cyberattacks cannot achieve this outcome. The potential bashes are not that devastating, the capabilities for cyber warfare are too widespread, and often the attacker cannot be conclusively identified. Deterrence policies of a different sort will have to develop. Libicki (2009) effectively analyzes the reasons why extant theories of deterrence, developed to prevent military aggression and the use of nuclear weapons, translate poorly to the case of cyberattacks.

Tor (2017) notes that the United States' deterrence policy, which has been critical to its US military strategy since WWII, is fundamentally different from that of modern-day Israel, where short bursts of violent attacks from enemies are accepted as a side effect of its deterrence strategy. There are strong parallels to criminal deterrence. Crime is never completely prevented. Its level is controlled, however, as would-be offenders are deterred by the criminal punishments they may face. Tor extrapolates from these examples to the deterrence of cyber attacks. He outlines a theory of "cumulative deterrence" where a series of attacks followed by bashing retaliations will keep constrain the level of attacks. Unlike threatened nuclear bashes, bashes in the cyber space will only impose moderate losses. Moreover, the bashes may not be responses in kind, as past bashes using lawsuits, tariffs and denials of technology illustrate. One proposed deterrence model for the United States government would allow enlisting private actors to "hack back" against cyber attackers on behalf of the government (Baker 2018).

Some analysts might classify a preventive attack on facilities intended to hack you as a deterrent measure: "We are seeking to prevent you from acquiring weapons that could harm us." The parallels to attacking bomb-building facilities are immediate. Legitimacy would of course be important. Thus, if there were international treaties against cyber warfare, such preventive strikes would enjoy greater legitimacy. They would also benefit if, as seems likely, the facilities had been used in past attacks. With cumulative deterrence, nations regularly classify their attacks as retaliatory efforts.

Computer hacks are almost always made anonymously. For example, non-state actors have been used as proxies in Russian cyberattacks against Ukraine (Maurer 2015). Even in non-proxied attacks, states may use "false flags" in malware design to conceal their identities (Skopik and Pahi 2020). It may be difficult for a state to justify a preventive attack if the

[2] To date, the most comprehensive international document discussing cyber warfare is the Tallinn Manual 2.0, a 2017 updated version of a manual created by a group of international experts shortly after a Tallinn, Estonia meeting of the NATO Cooperative Cyber Defence Centre of Excellence.

perceived threat cannot be confirmed to originate from the nation in question (Dipert 2010, 393). This may render preventive and retaliatory attacks more difficult to implement, especially in an environment with low tolerance for failures. However, Tor (2017, 100) argues that the attribution problem is often exaggerated, that the combination of logic and knowledge will usually identify the extremely likely guilty party. In a game-theoretic model of an attacker and defender able to retaliate with imperfect attribution, Welburn et al. (2019) discover several equilibria, none of which involves the defender perfectly signaling its retaliation capability. Instead, equilibria exist where no signaling occurs or where the defender imperfectly signals its capability to retaliate.

If the source cannot confidently be determined, as is sometimes the case, and if unjustified bashing would be significantly costly to the basher, deterrence will not work. Hackers frequently go to lengths to disguise their identity, for example by attacking through computers located in a third country or sometimes even through computers they have "captured" from others.

Disrupting in this manner is qualitatively different from the disruption in our model below. It sows confusion about the source of harm. In this analysis, the source is known. Disruption confuses about whether harm has taken place. Nevertheless, the results from below can be easily translated to the unknown source case.

Baliga, de Mesquita and Wolitzky (2019) (hereafter BMW) study the source-confusion case in cyberwarfare. There are n potential attackers. The potential basher who retaliates gains if the attacker is guilty; otherwise loses. They note an intriguing complementarity among attackers. More aggression by one, raises suspicions of him, and thereby protects others. That induces those others to become more aggressive. Nevertheless, a unique equilibrium is reached. Their model would apply broadly when one player is seeking to deter bad behavior from multiple potential guilty players. Think of police and known local thieves, or parents with children in a large family.

This analysis is closest to the contribution BMW make in their analyses of intelligence capabilities. Though straightforward quality improvements enhance deterrence, they show that other improvements in attribution can backfire and thereby weaken deterrence. The latter result, though arising in the source-confusion that underpins their paper, as opposed to action confusion in this analysis, is reminiscent of some counterintuitive results derived below.

To reiterate, and just to be clear, cyberattacks by private parties that have no state involvement are a significant problem. However, they are not our problem.

The analysis of the current paper, with its attention to deterrence that relies on intelligence and the potential for disruption, is broadly related to the literature on inspection games. That literature also began its life in an arms control context, but by now its applications have blossomed across an array of contexts. The general problem has an inspector verify whether agents have adhered to specified rules. Avenhaus et al. (2002) provides an extensive survey of this literature.

Inspection games are similar to our game in their sequence of moves. First, an agent decides whether or not to adhere to the rules. The agent then sends a signal, possibly purposefully noisy, of his action, The inspector observes the signal and decides whether or not to audit him, where auditing is an expensive or quantity constrained process. The analogy to our game is obvious. The agent is like our Player 1, and the inspector is like our Player 2. One important difference between our model and a typical inspection game is that in the latter, by auditing the agent, the inspector can detect with certainty whether or not he adhered to the rules. That determination is made before possibly taking tough measures against him. In our model Player 2's tough action of bashing Player 1 is taken under uncertainty since she can't detect with certainty what action Player 1 took. Finally, the signal in inspection games is strategically designed and sent by the agent. Here the signal is not Player 1's decision variable but rather the outcome of a machine owned by Player 2, but that machine is possibly disrupted by Player 1.

The general lesson from this literature review is that deterrence is a broadly observed phenomena. Effective deterrence almost always involves some form of monitoring, such as inspection or intelligence. The player being deterred, because he would like to get away playing his harmful action, will often seek to disrupt that monitoring. However, as the analysis below will show, it will often be detrimental to his interests to have that disruptive capability.

3 Model

1 moves first. 1 wants to take an action H that will benefit him, but will harm 2. 2 seeks to deter 1 by threatening to bash him if he has played H. 1's pure strategies are: H (take the harmful action), and NH (not take the harmful action). Player 2 can either bash, B, or not bash, NB. The following table describes the two players' payoffs for the four possible outcomes:

Table 2: Payoff Table

1 \ 2	NB	B
NH	$w_1, 1$	r_1, r_2
H	$1, 0$	$0, w_2$

It is assumed that $0 < r_i < w_i < 1,\ i = 1, 2$. That is, 1 ranks the outcomes (from best to worst) as follows: (H, NB), (NH, NB), (NH, B) and (H, B). 2 ranks the four possible outcomes (from best to worst) as follows: (NH, NB), (H, B), (NH, B) and (H, NB).

To determine whether or not to bash, 2 employs a noisy intelligence system to spy on Player 1 and to detect if he has taken a harmful action. The intelligence sends one of the two signals: h or nh, indicating imperfectly whether 1 takes a harmful action. The precision of intelligence is α, $\frac{1}{2} < \alpha < 1$. Namely, if 1 chooses either H or NH, then with probability α, the intelligence sends the signal h or nh, respectively. $\alpha = \frac{1}{2}$ means that the signal is sent completely randomly.

There is a disruptive technology that intercepts the signal with probability $\beta \in (0, 1)$. 1 may or may not have the technology. Whether he possesses it is 1's private information. 2 believes that 1 has the technology with probability g which, like all other parameters, is commonly known. That is, 1 can be one of two types. Type 1^+ possesses the disruptive technology; type 1^- does not have it.

If 1^+ chooses to operate the technology to Disrupt (D) the signal, then with probability β the accuracy of the signal drops to $\frac{1}{2}$ (a random signal is sent to 2). With probability $1 - \beta$, it fails to interfere with the signal, and the accuracy remains unchanged (namely, α). Therefore, in case of 1^+ choosing D, the precision of intelligence falls to $\alpha' = \frac{\beta}{2} + (1 - \beta)\alpha < \alpha$. If 1^+ chooses Not Disrupt (ND) or if the player is 1^-, the precision of intelligence remains at its original level, α.

The set of 1^+'s pure strategies is $S_{1^+} = \{(NH, D), (NH, ND), (H, D), (H, ND)\}$. The set of 1^-'s pure strategies is $S_{1^-} = \{H, NH\}$. The strategy (NH, D) indicates 1 operates his technology in an attempt to disrupt the signal sent to 2, but he does not harm 2, etc.

2, on the other hand, does not know whether the signal is disrupted. Based on the signal she receives, 2 decides whether or not (or with what probability) to bash 1.
The set of pure strategies for 2 is $S_2 = \{F, O, BB, NN\}$, which is defined as follows:

$$\begin{aligned}
F &\equiv B|h, NB|nh, \text{ follow the signal} \\
O &\equiv NB|h, B|nh, \text{ act opposite to the signal} \\
BB &\equiv B|h, B|nh, \text{ bash irrepective of the signal} \\
NN &\equiv NB|h, NB|nh, \text{ not bash irrepective of the signal}
\end{aligned}$$

Define the parameter set as

$$\mathcal{L} \equiv \left\{ \ell = (g, \alpha, \beta, r_1, w_1, r_2, w_2) | 0 \leq g \leq 1, \frac{1}{2} < \alpha < 1, 0 < \beta \leq 1, 0 < r_i < w_i < 1, i = 1, 2 \right\}$$

We assume that $\mathcal{L}$ is commonly known. The game Γ is described in the following game tree, and for $0 < g < 1$, it is a Bayesian game of asymmetric information.

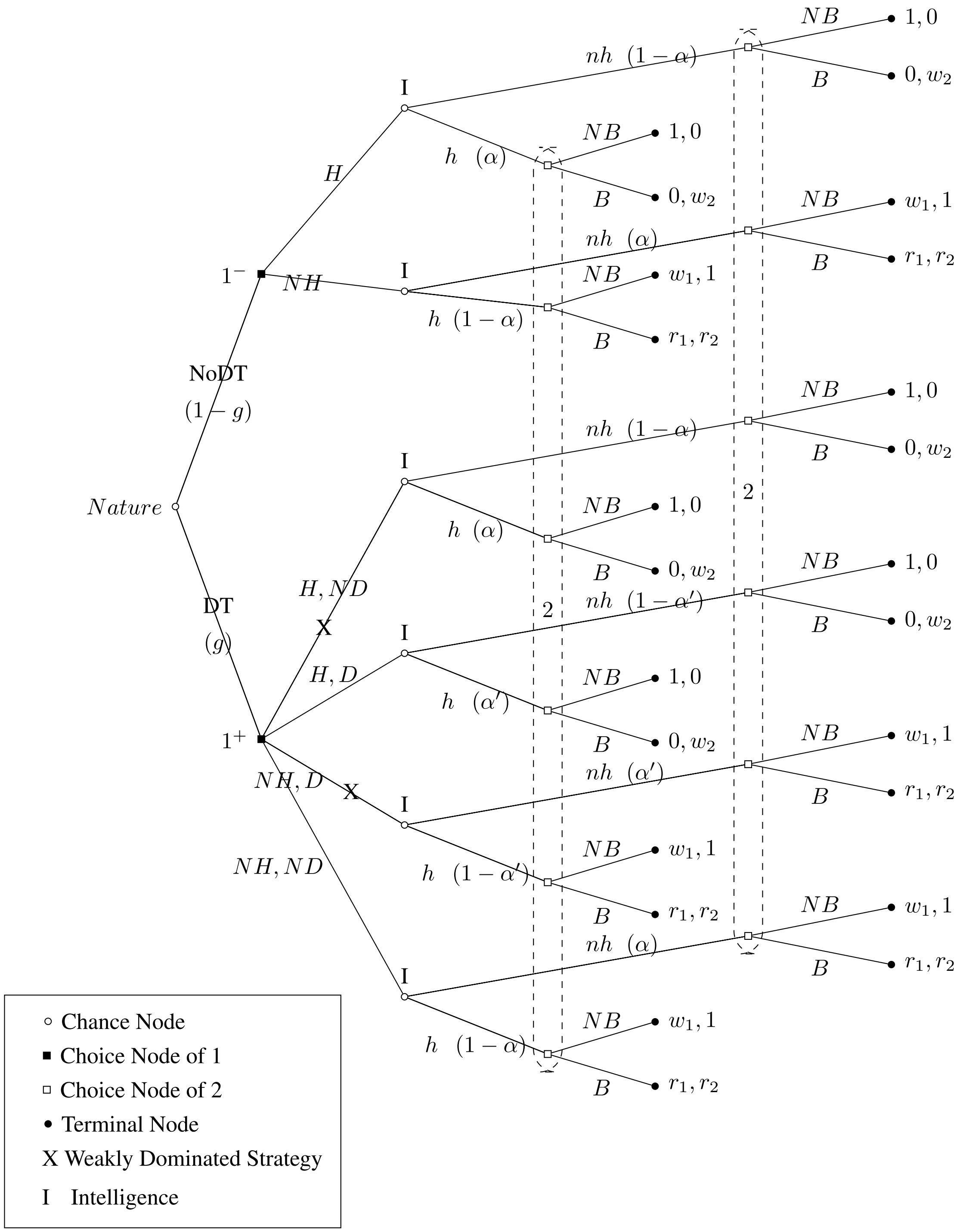

Figure 1: The Game Γ

3.1 The Analysis of the Asymmetric Information Case ($0 < g < 1$)

The asymmetric information case arises when 1 knows whether he has a technology for disruption, but 2 only knows the probability that he possesses it; hence, $0 < g < 1$.

Claim 1. *For every $\ell \in \mathcal{L}$,*

i. The strategy O of 2 is strictly dominated by F.

ii. In any equilibrium of Γ, 2 plays F with positive probability. She either plays pure F or mixes two of her pure strategies: F and NN or F and BB.

iii. 1^+ plays (H, ND) and (NH, D) with zero probability.

Proof: see the Appendix.

By Claim 1, acting opposite to the recommendation of the intelligence is a strictly dominated strategy for 2.

1^+ disrupts the signal only if he decides to take a harmful action. Otherwise, he is best off not disrupting the signal; precision helps him.

By Claim 1, for the equilibrium analysis of Γ, we can eliminate the strategies O of 2, and (NH, D) and (H, ND) of 1^+. Equivalently, we study the following reduced form game, Γ_r.

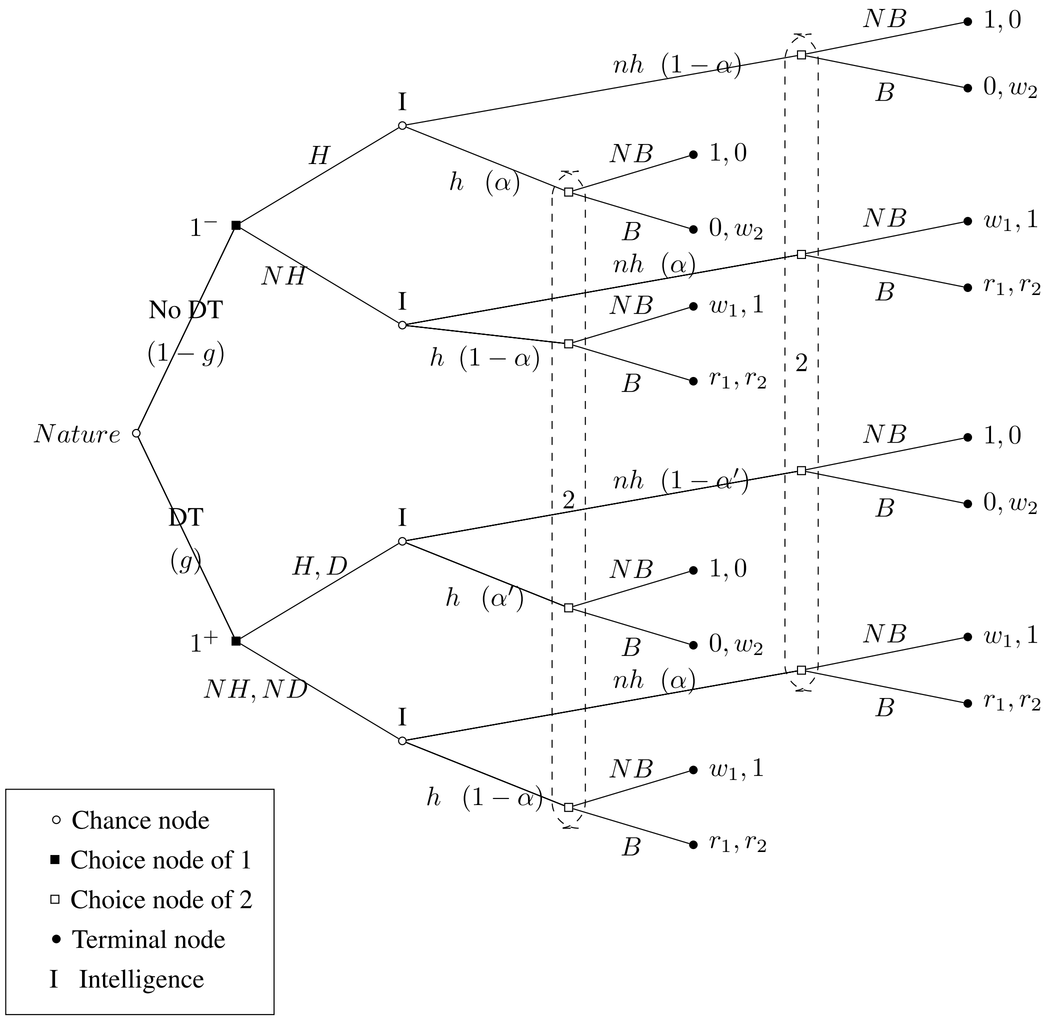

Figure 2: The Reduced Game Γ_r

To minimize notation, we denote the two active strategies (H, D) and (NH, ND) of 1^+, as H and NH, respectively. No confusion should arise, since he prefers combining D with H and ND with NH. We use the same notations H and NH for 1^-, who has no capability to disrupt.

The first proposition describes the equilibrium of Γ_r qualitatively. Its full characterization involves a partition of $\mathcal{L}$ into five regions and it is part of the proof of the proposition.

Proposition 1. *The game Γ_r has a unique equilibrium for each $\ell \in \mathcal{L}$. The equilibrium is continuous in ℓ and has the following properties:*

i. For every $\ell \in \mathcal{L}$, 1^+ places higher probability on H than 1^- does.

ii. A pure strategy equilibrium exists only in the region where $1 - w_1 > r_1$ and g and α take *intermediate values. In this equilibrium, 2 plays F, 1^+ plays H and 1^- plays NH.*

iii. *If g is relatively high and α is relatively low, 2 mixes F and BB. 1^+ mixes H and NH. 1^- plays pure NH.*

iv. *If g and α are both relatively low, 2 mixes F and BB. 1^+ plays pure H. 1^- mixes H and NH.*

v. *If α is high, 2 mixes F and NN. 1^+ mixes H and NH. 1^- plays pure NH.*

vi. *If g is relatively low and α is relatively high, 2 mixes F and NN. 1^+ plays pure H. 1^- mixes H and NH.*

Proof: see the Appendix.

For all parameters $\ell \in \mathcal{L}$, 1^+ always places higher probability on H than 1^- does. Namely, 1^+, who has a disruptive technology acts more aggressively than 1^- who does not. 1^+ never plays pure NH and 1^- never plays pure H. If 1^+ plays pure H, then 1^- mixes H and NH, or plays pure NH. If 1^+ mixes H and NH, 1^- plays pure NH.

By Proposition 1, a pure strategy equilibrium exists only when $1 - w_1 > r_1$, namely, when 1 cares more about safely harming 2 than avoiding getting caught by 2 and g and α take intermediate values. In this equilibrium, 1 takes the harmful action against 2 iff he has the disruptive technology, and 2 follows the signal and bashes 1 iff her intelligence sends the signal h. So with probability $1 - \alpha > 0$, 2 unjustifiably bashes an innocent 1. With the same probability 1^+, who harms 2, goes unpunished. The probability of these two mistakes of 2 diminish to zero as the precision of the intelligence approaches perfection.

Let us next turn to the mixed strategy equilibrium.

Suppose g is relatively high and α is relatively low. The intelligence of 2 has relatively low precision and 2 believes it is likely that 1 has a disruptive technology and 1 knows this. 2 thus can't give much credence to the signal she receives. Since her worst outcome (H, NB) is to allow 1 get unpunished if he harms her, she takes relatively aggressive action by mixing F and BB. This induces 1^+ to play less aggressively, and with positive probability he refrains from harming.

Regardless of g, if α is high, $1'$s action is reasonably likely to be correctly identified. Taking this into account 1 again acts cautiously. 1^+ mixes H and NH. 1^- plays pure NH. 2 in this region plays less aggressively than in the last region. Believing that it is likely 1 has taken the action NH, 2 trusts the signal nh, and trusts less the signal h. She mixes F and NN.

Consider next the region where g is relatively small and α takes relatively lower values. 2 believes 1^- is likely to be the actual type of 1. 1^+, contrary to the belief of 2, has a disruptive technology and is aware of the low quality of 2's intelligence. He uses his technology to further reduce the reliability of the signal and harms 2 with probability 1. 1^- who does not possess a disruptive technology acts a little less aggressively and he mixes H and NH. 2 reacts to the aggressive behavior of 1 accordingly, and she mixes BB with F.

As for the region where g is small and α is relatively high, the intuition is similar to that of the complete information case, where it is common knowledge that 1 does not possess a disruptive technology. Following Biran and Tauman (2009), in this case if α is relatively accurate, both players 1 and 2 act cautiously. 1^+, knowing his action will likely be identified, harms 2 with only a small probability. If 2 observes the signal h, she therefore trusts her intelligence less and bashes 1 with a relatively low probability. However, if she observes the likely signal nh, she chooses NB with certainty. This strategy is the result of a mixture of F and NN. The continuity of the equilibrium in ℓ extends this intuition to lower g and to high α.
We next analyze the equilibrium payoff of the players.

Proposition 2. *i. In a pure strategy equilibrium, the payoff of 1^+ is decreasing in α, while the payoffs of 1^- and 2 are both increasing in α.*

ii. In a mixed strategy equilibrium, the payoffs of 1^-, 1^+ and 2 are strictly increasing in α.

iii. In a pure strategy equilibrium, the payoff of 1^+ is increasing in β, the payoff of 1^- is constant in β, and 2's payoff is decreasing in β.

iv. In a mixed strategy equilibrium, 2's payoff is strictly decreasing in β.

If 1^- plays pure NH, the payoffs for both types of 1 are decreasing in β.

If 1^+ plays pure H, 1^+'s payoff is increasing in β and 1^-'s payoff is constant in β.

Proof: see the Appendix.
In $\mathcal{L}$, 2's payoff is increasing with the efficiency of her intelligence. Hence, her payoff is increasing in α and decreasing in β.

In a pure strategy equilibrium, a more efficient intelligence of 2, either due to high α or low β, lowers 1^+'s payoffs and raises 1^-'s payoff. The reason is that 1^+ who chooses H and 1^- who chooses NH have opposite interests. 1^+ wishes not to be detected, while 1^- wishes 2's intelligence to find him innocent. Therefore, 1^+ [1^-] is worse [better] off with high α and low β .

In a mixed strategy equilibrium, the strategies of 1^+ and 1^- have non-empty common support and the precision of 2's intelligence has some common impact on the payoffs of 1^+ and 1^-. Payoffs of both types of 1 are increasing in α. If g is high, the payoffs of 1^+ and 1^- are decreasing in β. If g is low, 1^+'s payoff is increasing while 1^-'s payoff is constant in β. As a result, 1^+ benefits from a better quality of his technology iff 2 believes it is more likely she faces 1^-, i.e., low g.

3.2 The Complete Information Case (g= 1, g=0)

In this section, we characterize the equilibrium of Γ_r when $g = 1$ or $g = 0$. That is, it is common knowledge that 1 has [does not have] a disruptive technology with certainty. Recall that $\alpha' = \frac{1}{2}\beta + \alpha(1-\beta)$.

Proposition 3. *(I.) Suppose $g = 1$. The game Γ_r has a unique equilibrium. 1 mixes H and NH. 2 either mixes F and BB or she mixes F and NN. Both players' payoffs are decreasing in β and increasing in α.*

i. If $\beta \geq 2w_1$ or $\alpha < \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$, 2 mixes F and BB. She follows the signal (F) with probability $q^* = \frac{r_1}{1-\alpha'-\alpha(w_1-r_1)}$, which is increasing in α and decreasing in β. 1 takes the harmful action (H) with probability $p^* = \frac{\alpha(1-r_2)}{(1-\alpha')w_2+\alpha(1-r_2)}$, which is increasing in α and decreasing in β.

ii. If $\beta < 2w_1$ and $\alpha > \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$, Player 2 mixes F and NN. She follows the signal (F) with probability $q^* = \frac{1-w_1}{\alpha'-(1-\alpha)(w_1-r_1)}$, which is decreasing in α and increasing in β. 1 takes the harmful action (H) with probability $p^* = \frac{(1-\alpha)(1-r_2)}{\alpha' w_2+(1-\alpha)(1-r_2)}$, which is decreasing in α and increasing in β.

(II.) Suppose $g = 0$. The equilibrium is the same as in (I.) with $\beta = 0$. Both players' payoffs are increasing in α.

(III.) Fixing all parameters but g, the equilibrium is continuous at $g = 1$ and $g = 0$.

Proof: see the Appendix.
We summarize the equilibrium outcome of Γ_r with $g = 1$ in Table 3.

Table 3: Properties of the Equilibrium of Γ_r when $g = 1$

	$\alpha < \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$ or $\beta \geq 2w_1$	$\alpha > \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$ and $\beta < 2w_1$
Equilibrium Region	Aggressive	Cautious
2's equilibrium strategy	$Pr_2(F) = q^* = \frac{r_1}{1-\alpha'-\alpha(w_1-r_1)}$, $Pr_2(BB) = 1 - q^*$	$Pr_2(F) = q^* = \frac{1-w_1}{\alpha'-(1-\alpha)(w_1-r_1)}$, $Pr_2(NN) = 1 - q^*$
1's equilibrium strategy	$p^* = \frac{\alpha(1-r_2)}{(1-\alpha')w_2+\alpha(1-r_2)}$	$p^* = \frac{(1-\alpha)(1-r_2)}{\alpha' w_2+(1-\alpha)(1-r_2)}$
2's payoff	$\Pi_2^* = w_2 \cdot \frac{(1-\alpha')r_2+\alpha(1-r_2)}{(1-\alpha')w_2+\alpha(1-r_2)}$	$\Pi_2^* = \frac{\alpha' w_2}{\alpha' w_2+(1-\alpha)(1-r_2)}$
1's payoff	$\Pi_1^* = \frac{r_1(1-\alpha')}{1-\alpha'-\alpha(w_1-r_1)}$	$\Pi_1^* = 1 - \frac{(1-w_1)\alpha'}{\alpha'-(1-\alpha)(w_1-r_1)}$
if α increases	$\Pi_2^* \uparrow$, $\Pi_1^* \uparrow$, $q^* \uparrow$, $p^* \downarrow$	$\Pi_2^* \uparrow$, $\Pi_1^* \uparrow$, $q^* \downarrow$, $p^* \uparrow$
if β increases	$\Pi_2^* \downarrow$, $\Pi_1^* \downarrow$, $q^* \downarrow$, $p^* \uparrow$	$\Pi_2^* \downarrow$, $\Pi_1^* \downarrow$, $q^* \uparrow$, $p^* \downarrow$

Note that $\frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta} \leq \frac{1}{2}$ iff $1-w_1 \leq r_1$, that is, 1 cares less about safely harming 2 than avoiding getting caught by 2 (see Table 2). In this case, the equilibrium is described in case ii of the proposition for all $\alpha \in (\frac{1}{2}, 1)$. 2 acts cautiously and either follows the signal or simply chooses to disregard the signal and not bash 1.

Suppose, by contrast, $1-r_2 > w_2$. Namely, the cost of 2 to bash an innocent player is higher than her cost of not punishing a harmful player. In this case, 2 acts cautiously and takes the harmful action with probability p^* less than $\frac{1}{2}$. The greater is the precision of the intelligence, the lower will be that probability.

The efficiency of 2's intelligence will depend on the conjunction of the precision (α) of her intelligence and the quality (β) of 1's disruptive technology. Efficiency changes positively in α and negatively in β. That efficiency partitions the equilibrium outcome into two regions, defined as "aggressive region" where intelligence is less efficient and "cautious region" where it is more efficient.

When both players know that 1's harmful action has a low probability of being detected, 1 takes the risk of choosing H with high probability. 2, thus fearful that 1 has taken the harmful action, acts more aggressively and mixes F (follow the signal) with BB (disregard the signal and bash 1).

When intelligence is efficient, both players pay cautiously. 1, knowing that his action is likely to be detected, plays H with only small probability. Deducing 1's strategy choice, 2 also acts cautiously. She mixes F (follow the signal) with NN (disregard the signal and not bash 1).

As intuition would suggest, a higher β has a similar effect to a smaller α, since either lowers the efficiency of 2's intelligence. That efficiency level matters, not the combination of α and β that produces it. In the aggressive [cautious] region, the probability p^* of H is increasing [decreasing] with the intelligence's efficiency. In the region where 2 mixes F and HH [F and NN], she places more [less] weight on playing BB (smaller q^*) [(greater q^*)].

Now let's turn to the impact of the intelligence on players' payoffs. When 2's intelligence is less efficient (either since the precision of the intelligence is low or the quality of disruptive technology is high), both players are worse off. In this case 2 acts more aggressively (see Proposition 1) which hurts 1 and hurts 2 as well because her bashing is more likely to be misdirected.

It was already noticed in BT and JTZ1 that greater precision α of the intelligence increases the payoff of 2 and quite surprisingly, also the payoff of 1. A more accurate intelligence deters 1 from harming (lower p^*). This makes 2 less concerned and lowers the probability she bashes. This in turn benefits 1 in addition to 2.

We just saw that 1's ability to deploy his disruptive technology hurts him as well as 2, and they both are hurt more the better is that technology. A natural question then arises: why doesn't 1, whose payoff is decreasing in the quality of his disruptive technology, simply turn it off? Both he and 2 would gain from a convincing turn off. The unfortunate reality is that he could not make a "turn off" action credible. That is because if 2 believed that it was turned off, 1 would be better to keep it turned on. The situation is shown in Table 4.

Table 4: The Impossibility of Turning Off Disruption

1 chooses \ 2 believes	Technology On	Technology Off
Technology On	3rd, 2nd	1st, 3rd
Technology Off	not ranked, not ranked	2nd, 1st

Both players prefer the lower-right box to the upper-left box, as we observed. But the latter is not an equilibrium, since if 2 believed Off 1 would choose On. Moreover, a mixed strategy for 1, if he did possess the disruptive technology, would also not be credible. He would have the incentive to deviate to a pure On strategy.

<u>Observation</u>

i. Suppose g is sufficiently small. Then 1^+'s payoff is increasing in β. 1^-'s payoff is constant in β.

ii. Suppose g is sufficiently large. Then 1^+'s and 1^-'s payoffs are both decreasing in β.

That is, if 2 strongly believes that 1 has no disruptive technology, then possessing such a technology benefits him. If 2, however, strongly believes 1 possesses it, 1's payoff decreases with the quality of this technology whether or not he actually possesses it.

3.3 Example

The following numerical example demonstrates that a better intelligence induces both players to be more cautious and both increase their payoff. A stronger belief of 1 possessing a disruptive technology makes 2 more aggressive, and 1 behaves more cautiously.

Table 5: Numerical Example of Game Γ_r

	1's equilibrium strategy	2's equilibrium strategy
$r_1 = 0.2, r_2 = 0.3,$ $w_1 = 0.5, w_2 = 0.8,$ $\alpha = 0.6, \beta = 0.5, g = 0.5$	1^+ for sure harms 1^- harms with probability 0.081 $\Pi^*_{1+} = 0.409, \ \Pi^*_{1-} = 0.364$	2 chooses F with probability 0.910 and she chooses BB with probability 0.090 $\Pi^*_2 = 0.570$
$r_1 = 0.2, r_2 = 0.3,$ $w_1 = 0.5, w_2 = 0.8,$ $\alpha = 0.9, \beta = 0.5, g = 0.5$	1^+ harms with probability 0.222 1^- does not harm $\Pi^*_{1+} = \Pi^*_{1-} = 0.478$	2 chooses F with probability 0.746 and she chooses NN with probability 0.254 $\Pi^*_2 = 0.889$
$r_1 = 0.2, r_2 = 0.3,$ $w_1 = 0.5, w_2 = 0.8,$ $\alpha = 0.6, \beta = 0.5, g = 0.95$	1^+ harms with probability 0.567 1^- does not harm $\Pi^*_{1+} = \Pi^*_{1-} = 0.333$	2 chooses F with probability 0.741 and she chooses BB with probability 0.259 $\Pi^*_2 = 0.569$

4 Extensions

4.1 The Case when a tampered signal is always detected

One extension is to add an uninformative signal that reveals itself, say by coming across as static or with an identifiable marker. That is, a disruption, which is successful with probability β, causes the intelligence to send an uninformative signal, t (stands for tampered). Upon receiving a tampered signal, 2 knows for sure that 1 has a disruptive technology, has used it, and that disruption has been successful. The properties of tampering are common knowledge. The extended game tree is described in Figure 3.

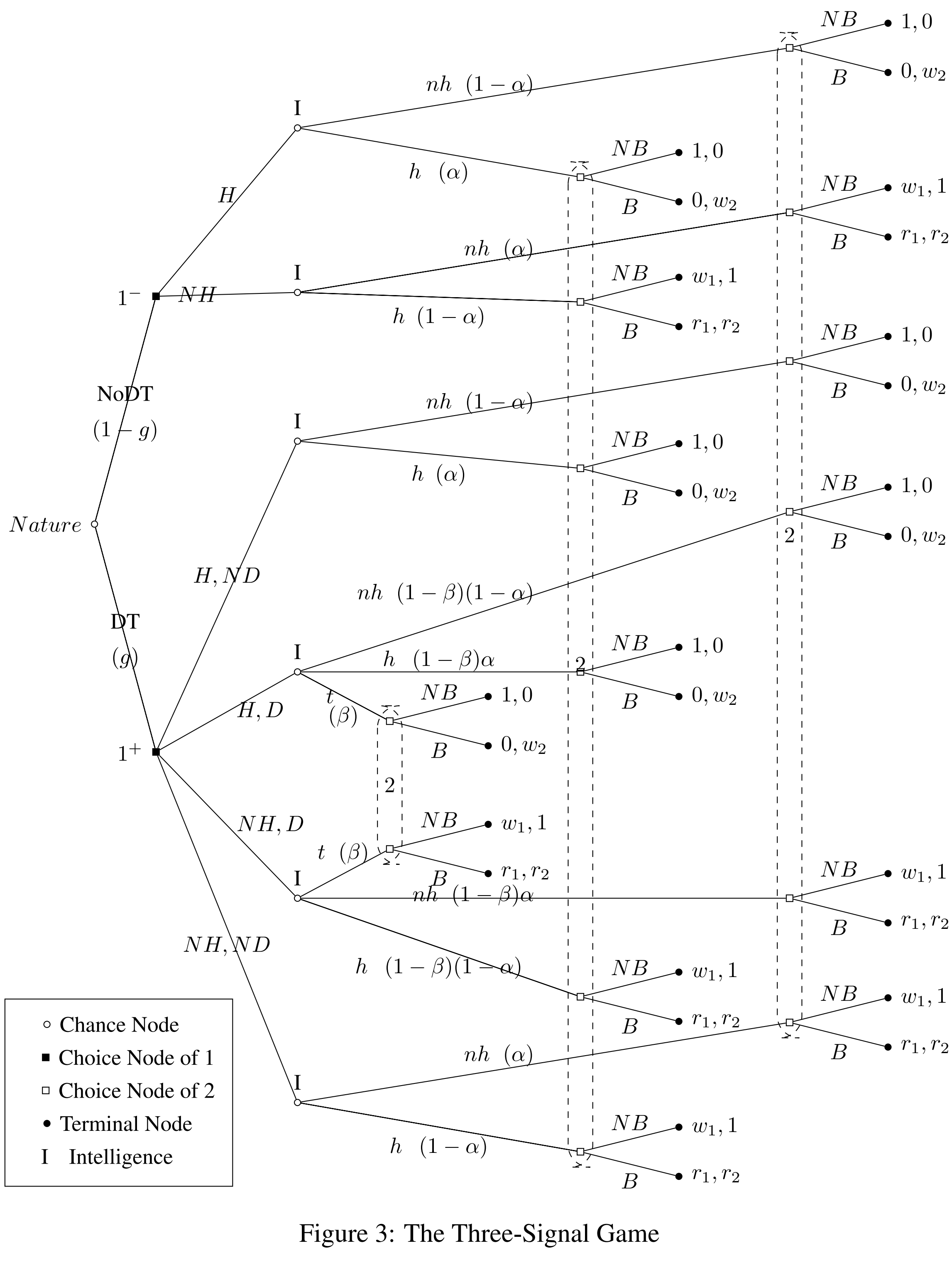

Figure 3: The Three-Signal Game

In this game, 2 may receive one of the three signals (s): a tampered signal (t), a harmful signal

h or an unharmful signal nh. Namely, $s \in \{t, h, nh\}$. The quality of the disruptive technology remains β. That is, the probability of producing a signal t is β provided that 1^+ chooses to operate his technology.

4.2 Preemptive and After-the-Fact Deterrence

Another extension would consider a preemptive deterrence action by 2. Before 1 decides H or NH, 2 makes a investment to build her bash and intelligence capability.

The market-grabbing illustration builds on an elegant literature on entry deterrence. There are two prime players, the incumbent and the potential entrant, though in some cases third parties are involved to help the incumbent protect her position. The classic formulation (Tirole 1988) has the incumbent move first. Her move might involve fierce advertising, building excess capacity, or licensing her technology to others as a means to signal to the potential entrant that his entry will not be profitable. Such models might be labelled preemptive deterrence. A second class of deterrence models, best known from military affairs, involves after-the-fact deterrence. The underpinnings of that model were worked out in the late 1950s (Schelling 1960). In the decades of the Cold War, after-the-fact deterrence enabled the United States and the Soviet Union to deter each other from launching a nuclear attack or indeed a more general military attack, with their bashing threats of retaliation. In recent years, the most salient example is a nation 1 that would like to build a nuclear weapon. Nation 2, is known to have the capability to wipe out, i.e., bash, 1's facilities. Our third example, industrial espionage, also involves after-the-fact deterrence. If a nation, or companies within that nation proceed to steal intellectual property that nation will be bashed with tariffs. In our entry deterrence, bomb building and industrial espionage examples, but not the Cold War example only mentioned in this literature review, 1 will take his move in secret. 2 will seek to determine that move, because she suffers if she bashes without justification or fails to bash when 1 has taken the harmful action (taking steps to enter, building the bomb, stealing the intellectual property). Her means to do so involves operating an intelligence system. How the two players choose their actions in the presence of such a system, including 1's potential to disrupt it, are the subject of this paper.

4.3 Probabilistic Bashing

Our formulation always has 1 always face a bash threat that is probabilistic. That is, he never takes a harmful move that will surely entail bashing. In that way, it is reminiscent of Schelling's (1960, pp. 187-206) famed "The Threat That Leaves Something to Chance". The motivation for making the threat merely probabilistic differs between our analysis and Schelling's. Schelling was concerned with maintaining the credibility of a threat that was too monumental, e.g., launching a nuclear weapon. If the threatened action is too costly, credibility would be lost. Making it probabilistic scales down its expected cost. Schelling also makes it clear that 2 must relinquish control on whether to carry out the bashing action for three reasons: Otherwise, 1 will engage in a guessing game of what 2 will do; the cost of bashing for 2 would be too high for her to carry through on her threat; 2's credibility for future encounters will be lost if 1 harms, and she is in control but fails to bash. Our bash threat is probabilistic because 2 faces inevitable uncertainty when assessing what 1 has done. Hence, 2 must employ a mixed strategy to best deter 1 while minimizing the expected costs of not bashing when she should and bashing when she should not. In the market grab context, 1 proceeds to design-around iff and only if he believes that his efforts will go undetected. 2 will launch her own expensive R&D effort to secure an improvement patent if she thinks it sufficiently likely that 1 has chosen the design-around strategy. 2 will bash 1's facilities if she thinks it sufficiently likely that he is building a bomb; etc.

The initial two sections, (3.1) and (3.2), of our analysis are intended to lay out the role of intelligence and its disruption in a deterrence game. They involve two actions, where 1 goes first. The third section addresses three-action situations. First, 2 establishes his bashing threat through some combination of an intelligence system and bash capability. Second, 1 takes his action. Third, 2, given her intelligence, chooses whether to bash. This third section shows how to add a third step to classic preemptive deterrence and after-the-fact deterrence models.

5 Concluding Thoughts

Deterrence models have played a significant role in the academic literature since the late 1950s. They were initially inspired by the need to deter nuclear attacks and military aggression more generally, but have recently been applied in many other contexts. Of course, deterrence as a practice has played a role for eons, for example when a parent employed it to prevent a child from misbehaving. Intelligence systems played a role in long-ago deterrence as they do today. For example, they helped determine whether a child really hit a sibling or slacked off when charged to hunt for berries. Presumably children, in the spirit of this paper, attempted to disrupt intelligence. For example, they would only hit when the parent was thought to be out of sight and outside earshot. They would slack off when berries were abundant, and they had picked a quantity that would normally suffice. We have shown that deterrence is a strategy that is employed broadly. When it is, potential harmers will seek to choose in secret. Potential bashers will therefore need to employ an intelligence system lest they bash without justification. Potential harmers may counter that system with a disruptive technology. Surprisingly, employing that technology may hurt rather than help them. Yet, potential harmers may be unable to foreswear the use of a technology that actually hurts them. In short, the complex game of deterrence has many twists and turns. Analytic models can help us see what is around the next corner.

6 Bibliography

1. Avenhaus, Rudolf, Bernhard Von Stengel, and Shmuel Zamir. "Inspection games." Handbook of game theory with economic applications 3 (2002): 1947-1987.

2. Baker, Benjamin. "Considering the Potential Deterrence Value of Legislation Allowing Hacking Back." SSRN Scholarly Paper ID 3319530. (2018) Rochester, NY: Social Science Research Network.

3. Baliga, Sandeep, and Tomas Sjostrom. "Strategic ambiguity and arms proliferation." Journal of Political Economy 116 (2008): 1023-1057.

4. Baliga, S., de Mesquita, E.B. and Wolitzky, A.. "Deterrence with imperfect attribution." (2019) Working Paper.

5. Barrachina, Alex, Yair Tauman, Amparo Urbano. "Entry and espionage with noisy signals." Games and Economic Behavior 83 (2014): 127-146.

6. Bas, Muhammet A., and Andrew J. Coe. "Arms diffusion and war." Journal of Conflict Resolution 56 (2012): 651-674.

7. Biran Dov and Yair Tauman, "The Decision to Attack a Nuclear Facility: The Roll of Intelligence" (2009) Unpublished manuscript.

8. Chen, Thomas M. 2010. "Stuxnet, the Real Start of Cyber Warfare? [Editor's Note]." IEEE Network 24 (6): 2–3.

9. Cioffi-Revilla, Claudio. "Origins and age of deterrence: Comparative research on old world and new world systems." Cross-Cultural Research 33 (1999): 239–64.

10. Coe, Andrew J., and Jane Vaynman. "Why arms control is so rare." American Political Science Review 114 (2020): 342-355.

11. Cohen, W.M., Nelson, R.R. and Walsh, J.P., . "Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not)" (2000) National Bureau of Economic Research (No. w7552).

12. Debs, Alexandre, and Nuno P. Monteiro. "Known unknowns: Power shifts, uncertainty, and war." International Organization 68 (2014): 1-31.

13. Dipert, Randall R. 2010. "The Ethics of Cyberwarfare." Journal of Military Ethics 9 (4): 384–410.

14. Duelfer, Charles A., and Stephen Benedict Dyson. "Chronic misperception and international conflict: The U.S.-Iraq experience." International Security 36 (2011): 73-100.

15. Ellison, Glenn, and Sara Fisher Ellison. 2011. “Strategic Entry Deterrence and the Behavior of Pharmaceutical Incumbents Prior to Patent Expiration.” American Economic Journal: Microeconomics 3 (1): 1–36.

16. Eswaran, Mukesh. “Licensees as Entry Barriers.” The Canadian Journal of Economics / Revue Canadienne D’Economique 27 (1994): 673-88.

17. Fearon, James D. “Rationalist explanations for war.” International Organization 49, no. 3 (1995): 379-414.

18. Gallini, Nancy T. “Deterrence by market sharing: A strategic incentive for licensing.” The American Economic Review 74 (1984): 931-941.

19. Hall, B.H. and Ziedonis, R.H.. “The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1979-1995.” Rand Journal of Economics (2001): 101-128.

20. Harrington, Joseph E.. “Limit Pricing When the Potential Entrant Is Uncertain of Its Cost Function: 1. Introduction.” Econometrica (1986-1998); Evanston 54 (2): 429.

21. Henriksen, Anders. “The End of the Road for the UN GGE Process: The Future Regulation of Cyberspace.” Journal of Cybersecurity (2019): 5 (1).

22. Jelnov Artyom, Yair Tauman, and Richard Zeckhauser. “Attacking the unknown weapons of a potential bomb builder: The impact of intelligence on the strategic interaction.” Games and Economic Behavior 104 (2017): 177-189.

23. Jelnov Artyom, Yair Tauman, and Richard Zeckhauser. “Confronting an enemy with unknown preferences: Deterrer or provocateur?.” European Journal of Political Economy 54 (2018): 124-143.

24. Klemperer, Paul. “Entry Deterrence in Markets with Consumer Switching Costs.” The Economic Journal (1987) 97: 99–117.

25. Langinier, C.. “Are patents strategic barriers to entry?” Journal of Economics and Business 56(5), (2004):349-361.

26. O’Neill, Barry. “Game theory models of peace and war.” In: R.J. Aumann and S. Hart (eds.), Handbook of Game Theory with Economic Applications, vol. 2. Amsterdam: North Holland (1994): 995-1053.

27. Lee, Robert M., Michael J. Assante, and Tim Conway. “Analysis of the Cyber Attack on the Ukrainian Power Grid.” SANS Industrial Control Center, Electricity Information Sharing and Analysis Center. (2016)

28. Libicki, Martin C. Cyberdeterrence and Cyberwar. Rand Corporation. (2009)

29. Maurer, Tim. "Cyber Proxies and the Crisis in Ukraine." In Cyber Warfare in Perspective: Russian Aggression Against Ukraine (2015) 79–86. Tallinn, Estonia: NATO Cooperative Center of Cyber Defence.

30. Milgrom, Paul, and John Roberts. "Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis." Econometrica (1982) 50 (2): 443–59.

31. Perlroth, Nicole, and Clifford Krauss. "A Cyberattack in Saudi Arabia Had a Deadly Goal. Experts Fear Another Try." The New York Times, (March 15, 2018) sec. Technology.

32. Powell, Robert. "Nuclear Deterrence Theory: The Search for Credibility." Cambridge, UK: Cambridge University Press (1990).

33. Qian, Yi. "Impacts of Entry by Counterfeiters." The Quarterly Journal of Economics (2008) 123 (4): 1577–1609.

34. Rockett, K.E. "Choosing the competition and patent licensing." The RAND Journal of Economics (1990): 161-171.

35. Sanger, David E. "Obama Order Sped Up Wave of Cyberattacks Against Iran." The New York Times (June 1, 2012) sec. World.

36. Schelling, Thomas C. "The Strategy of Conflict." Cambridge, MA: Harvard University Press (1980 edition) (initial edition 1960).

37. Schelling, Thomas C. "An Astonishing Sixty Years: The Legacy of Hiroshima." Nobel Prize Lecture (2005, December 8).

38. Schmitt, Michael N, and NATO Cooperative Cyber Defence Center of Excellence. "Tallinn Manual 2.0: On the International Law Applicable to Cyber Operations." (2017) 2nd ed. Cambridge, United Kingdom: Cambridge University Press.

39. Skopik, Florian, and Timea Pahi. "Under False Flag: Using Technical Artifacts for Cyber Attack Attribution." Cybersecurity (2020) 3 (1): 1–20.

40. Slowik, Joe. "Anatomy of an Attack: Detecting and Defeating CRASHOVERRIDE." Dragos Industrial Cybersecurity (2018).

41. Stevens, Clare. "Assembling Cybersecurity: The Politics and Materiality of Technical Malware Reports and the Case of Stuxnet." Contemporary Security Policy (2020) 41 (1): 129–52.

42. Tor, Uri. " 'Cumulative Deterrence' as a New Paradigm for Cyber Deterrence." Journal of Strategic Studies (2017) 40 (1–2): 92–117.

43. Warrick, Joby. "Foreign Intelligence Officials Say Attempted Cyberattack on Israeli Water Utilities Linked to Iran." Washington Post (May 8, 2020).

44. Welburn, Jonathan William, Justin Grana, and Karen Schwindt. "Cyber Deterrence or: How We Learned to Stop Worrying and Love the Signal." Product Page. RAND Corporation (2019).

7 Appendix

Proof of Claim 1. The associated game Γ in strategic form is described below.

Table 6: Γ in Strategic Form

$1^+, 1^-$ \ 2	NN	F	O	BB
$(NH, ND), NH$	$w_1, w_1, 1$	$\alpha w_1 + (1-\alpha)r_1,$ $\alpha w_1 + (1-\alpha)r_1,$ $\alpha + (1-\alpha)r_2$	$(1-\alpha)w_1 + \alpha r_1,$ $(1-\alpha)w_1 + \alpha r_1,$ $1 - \alpha + \alpha r_2$	r_1, r_1, r_2
$(NH, D), NH$	$w_1, w_1, 1$	$\alpha' w_1 + (1-\alpha')r_1,$ $\alpha w_1 + (1-\alpha)r_1,$ $g[\alpha' + (1-\alpha')r_2] + (1-g)[\alpha + (1-\alpha)r_2]$	$(1-\alpha')w_1 + \alpha' r_1,$ $(1-\alpha)w_1 + \alpha r_1,$ $g(1-\alpha' + \alpha' r_2) + (1-g)(1-\alpha+\alpha r_2)$	r_1, r_1, r_2
$(NH, ND), H$	$w_1, 1, g$	$\alpha w_1 + (1-\alpha)r_1,$ $1-\alpha,$ $g[\alpha + (1-\alpha)r_2] + (1-g)\alpha w_2$	$(1-\alpha)w_1 + \alpha r_1,$ $\alpha,$ $g(1-\alpha+\alpha r_2) + (1-g)(1-\alpha)w_2$	$r_1, 0, gr_2 + (1-g)w_2$
$(NH, D), H$	$w_1, 1, g$	$\alpha' w_1 + (1-\alpha')r_1,$ $1-\alpha,$ $g[\alpha' + (1-\alpha')r_2] + (1-g)\alpha w_2$	$(1-\alpha')w_1 + \alpha' r_1,$ $\alpha,$ $g(1-\alpha' + \alpha' r_2) + (1-g)(1-\alpha)w_2$	$r_1, 0, gr_2 + (1-g)w_2$
$(H, D), NH$	$1, w_1, 1-g$	$1-\alpha',$ $\alpha w_1 + (1-\alpha)r_1,$ $g\alpha' w_2 + (1-g)[\alpha + (1-\alpha)r_2]$	$\alpha',$ $(1-\alpha)w_1 + \alpha r_1,$ $g(1-\alpha')w_2 + (1-g)(1-\alpha+\alpha r_2)$	$0, r_1, gw_2 + (1-g)r_2$
$(H, ND), NH$	$1, w_1, 1-g$	$1-\alpha,$ $\alpha w_1 + (1-\alpha)r_1,$ $g\alpha w_2 + (1-g)[\alpha + (1-\alpha)r_2]$	$\alpha,$ $(1-\alpha)w_1 + \alpha r_1,$ $g(1-\alpha)w_2 + (1-g)(1-\alpha+\alpha r_2)$	$0, r_1, gw_2 + (1-g)r_2$
$(H, D), H$	$1, 1, 0$	$1-\alpha',$ $1-\alpha,$ $g\alpha' w_2 + (1-g)\alpha w_2$	$\alpha',$ $\alpha,$ $g(1-\alpha')w_2 + (1-g)(1-\alpha)w_2$	$0, 0, w_2$
$(H, ND), H$	$1, 1, 0$	$1-\alpha,$ $1-\alpha,$ $g\alpha w_2 + (1-g)\alpha w_2$	$\alpha,$ $\alpha,$ $g(1-\alpha)w_2 + (1-g)(1-\alpha)w_2$	$0, 0, w_2$

(i) Since $\frac{1}{2} < \alpha' = \frac{\beta}{2} + (1-\beta)\alpha < \alpha$, it is easy to see that O is dominated by F for 2.

(ii) Clearly, no equilibrium exists in which 2 plays pure NN or pure BB. Next we prove 2 does not mix NN and BB.

After observing the signal sent by the intelligence, 2's conditional probabilities over 1's actions are denoted by $P_2(H|h), P_2(NH|h) = 1 - P_2(H|h), P_2(H|nh)$ and $P_2(NH|nh) = 1 - P_2(H|nh)$.

2's expected payoffs conditioned on the signal are

$$\Pi_2(B|h) = P_2(H|h)w_2 + P_2(NH|h)r_2, \qquad \Pi_2(NB|h) = P_2(NH|h)$$
$$\Pi_2(B|nh) = P_2(H|nh)w_2 + P_2(NH|nh)r_2, \quad \Pi_2(NB|nh) = P_2(NH|nh)$$

Then we have $B|h \succeq_2 NB|h \longleftrightarrow P_2(H|h) \geq \frac{1-r_2}{1-r_2+w_2}$.

Similarly, $B|nh \succeq_2 NB|nh \longleftrightarrow P_2(H|nh) \geq \frac{1-r_2}{1-r_2+w_2}$.

Then 2 is best off playing

$$\begin{cases} NN, & \text{if } P_2(H|h) < \frac{1-r_2}{1-r_2+w_2} \text{ and } P_2(H|nh) < \frac{1-r_2}{1-r_2+w_2} \\ BB, & \text{if } P_2(H|h) > \frac{1-r_2}{1-r_2+w_2} \text{ and } P_2(H|nh) > \frac{1-r_2}{1-r_2+w_2} \end{cases}$$

Suppose 2 is indifferent between NN and BB. Then $P_2(H|h) = P_2(H|nh) = \frac{1-r_2}{1-r_2+w_2}$. This implies that the signal 2 receives is uninformative, contradicting $\alpha > \alpha' > \frac{1}{2}$. Formally,

$$P_2(H|h) = \frac{P_2(h|H)Pr(H)}{P_2(h|H)Pr(H) + P_2(h|NH)Pr(NH)} \tag{1}$$

$$P_2(H|nh) = \frac{P_2(nh|H)Pr(H)}{P_2(nh|H)Pr(H) + P_2(nh|NH)Pr(NH)} \tag{2}$$

where $Pr(H)$ is the probability that 1 (either type) takes the harmful action. $P_2(h|H)$ and $P_2(nh|NH)$ measure the expected precision of the signal, which is in between α and α' and greater than $\frac{1}{2}$. Since $P_2(H|h) = P_2(H|nh)$, we have, by (1) and (2),
$2\alpha' < P_2(h|H) + P_2(nh|NH) = 1$, a contradiction to $\alpha' > \frac{1}{2}$.

Hence in any equilibrium of Γ, 2 must place positive probability on F.
(iii) Since 2 plays F with positive probability and $\alpha' < \alpha$, the expected payoff to 1^+ by playing (NH, ND) is greater than that by playing (NH, D). Similarly, the expected payoff to him by playing (H, D) is greater than that by playing (H, ND). Hence in equilibrium, 1^+ must assign zero probability to both (NH, D) and (H, ND).

□

Proof of Proposition 1 and 2. Consider the reduced game Γ_r. Clearly, there is no pure strategy equilibrium where 2 plays BB or NN (see Claim 1).

Suppose 2 plays pure F. 1^+ obtains $\alpha w_1 + (1-\alpha)r_1$ by playing NH, and $1-\alpha'$ by playing H. He strictly prefers NH iff $\alpha > \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$. 1^- obtains $\alpha w_1 + (1-\alpha)r_1$ by playing NH, and $1-\alpha$ by playing H. He strictly prefers NH iff $\alpha > \frac{1-r_1}{1-r_1+w_1}$. Hence 2 is better off deviating to NN if $\alpha > \max\{\frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}, \frac{1-r_1}{1-r_1+w_1}\}$ and she is better off deviating to BB if $\alpha < \min\{\frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}, \frac{1-r_1}{1-r_1+w_1}\}$.

So if a pure strategy equilibrium exists, then either $\frac{1-r_1}{1-r_1+w_1} \leq \alpha \leq \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$ or $\frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta} \leq \alpha \leq \frac{1-r_1}{1-r_1+w_1}$. Note that if the latter is true, then $\frac{1-r_1}{1-r_1+w_1} \geq \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$ and this is equivalent to $1-w_1 \leq r_1$, but in this case $\frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta} \leq \frac{1-r_1}{1-r_1+w_1} \leq \frac{1}{2} < \alpha$, a contradiction. Hence $1-w_1 > r_1$ must hold and a pure strategy equilibrium may exist only if

$\frac{1-r_1}{1-r_1+w_1} < \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$ and

$$\frac{1-r_1}{1-r_1+w_1} \leq \alpha \leq \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta} \tag{3}$$

Suppose first that (3) holds with strict inequalities. Then 1^+ chooses H, 1^- chooses NH and 2 chooses F. To guarantee that 2 has no incentive to deviate from F in this case, the inequality

$$(1-g)[\alpha+(1-\alpha)r_2]+g\alpha' w_2 \geq \max\{gw_2+(1-g)r_2, 1-g\}$$

must hold. Equivalently,

$$\alpha \geq \max\{\frac{(1-\frac{1}{2}\beta)gw_2}{(1-\beta)gw_2+(1-g)(1-r_2)}, 1-\frac{(1-\frac{1}{2}\beta)gw_2}{(1-\beta)gw_2+(1-g)(1-r_2)}\} \tag{4}$$

Combining (3) and (4), we have

$$\max\{\frac{(1-\frac{1}{2}\beta)gw_2}{(1-\beta)gw_2+(1-g)(1-r_2)}, 1-\frac{(1-\frac{1}{2}\beta)gw_2}{(1-\beta)gw_2+(1-g)(1-r_2)}, \frac{1-r_1}{1-r_1+w_1}\} < \alpha < \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta} \tag{5}$$

Next suppose one side of (3) holds as an equality. If $\alpha = \frac{1-r_1}{1-r_1+w_1}$, then 1^- is indifferent between H and NH, and 1^+ strictly prefers H. In this case, 2 is better off deviating from F to BB. If $\alpha = \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$, then 1^+ is indifferent between H and NH, and 1^- strictly prefers NH. In this case, 2 is better off deviating from F to NN.

The set of α, $\frac{1}{2} < \alpha < 1$, satisfying (5) is non-empty iff $\frac{(1-\frac{1}{2}\beta)gw_2}{(1-\beta)gw_2+(1-g)(1-r_2)} < \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$ and $1-\frac{(1-\frac{1}{2}\beta)gw_2}{(1-\beta)gw_2+(1-g)(1-r_2)} < \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$. Equivalently,

$\frac{(1-r_2)(w_1-\frac{1}{2}\beta)}{(1-r_2)(w_1-\frac{1}{2}\beta)+w_2[1-r_1-\beta+(w_1+r_1)\frac{1}{2}\beta]} < g < \frac{(1-r_2)(1-r_1-\frac{1}{2}\beta)}{(1-r_2)(1-r_1-\frac{1}{2}\beta)+w_2[w_1-(w_1+r_1)\frac{1}{2}\beta]}$.

It can be verified that the subset of the parameter set $\mathcal{L}$ satisfying these constraints is non empty iff $1-w_1 > r_1$ and $0 < \beta \leq 2w_1$.

The pure strategy equilibrium payoff of 1^+ is $1-\alpha'$ and it is increasing in β and decreasing in α. The equilibrium payoff of 1^- is $\alpha w_1+(1-\alpha)r_1$ and it is increasing in α and independent of β. The equilibrium payoff of 2 is $g\alpha' w_2+(1-g)[\alpha+(1-\alpha)r_2]$ and it is increasing in α and decreasing in β.
We next analyze a mixed strategy equilibrium.

Suppose 1^+ chooses H with probability p_+, and 1^- chooses H with probability p_-. Conditioning on the signal, 2 assigns probabilities over 1's actions (see Figure 2):

$$P_2(H|h) = \frac{g\,p_+\,\alpha' + (1-g)\,p_-\,\alpha}{g(1-p_+)(1-\alpha) + gp_+\alpha' + (1-g)(1-p_-)(1-\alpha) + (1-g)p_-\alpha},$$
$$P_2(NH|h) = 1 - P_2(H|h),$$
$$P_2(H|nh) = \frac{g\,p_+(1-\alpha') + (1-g)p_-(1-\alpha)}{g(1-p_+)\alpha + gp_+(1-\alpha') + (1-g)(1-p_-)\alpha + (1-g)p_-(1-\alpha)},$$
$$\text{and} \quad P_2(NH|nh) = 1 - P_2(H|nh),$$

where $\alpha' = \frac{\beta}{2} + (1-\beta)\alpha$. 2's conditional expected payoffs are

$$\Pi_2(B|h) = P_2(H|h)w_2 + P_2(NH|h)r_2, \qquad \Pi_2(NB|h) = P_2(NH|h)$$
$$\Pi_2(B|nh) = P_2(H|nh)w_2 + P_2(NH|nh)r_2, \quad \Pi_2(NB|nh) = P_2(NH|nh)$$

Given (p_+, p_-), for 2,

$$B|h \succ_2 NB|h \longleftrightarrow P_2(H|h) > \frac{1-r_2}{1-r_2+w_2}, \;\longleftrightarrow\; \alpha > \gamma(p_+, p_-) \tag{6}$$

where

$$\gamma(p_+, p_-) \equiv \frac{(1-r_2)[g(1-p_+) + (1-g)(1-p_-)] - \frac{1}{2}\beta g p_+ w_2}{(1-r_2)[g(1-p_+) + (1-g)(1-p_-)] + w_2[gp_+(1-\beta) + (1-g)p_-]} \tag{7}$$

Namely, after observing the signal h, 2 strictly prefers B to NB iff $\alpha > \gamma(p_+, p_-)$, strictly prefers NB to B iff $\alpha < \gamma(p_+, p_-)$ and is indifferent with NB and B iff $\alpha = \gamma(p_+, p_-)$.

Similarly,

$$B|nh \succ_2 NB|nh \longleftrightarrow P_2(H|nh) > \frac{1-r_2}{1-r_2+w_2}, \;\longleftrightarrow\; \alpha < 1 - \gamma(p_+, p_-) \tag{8}$$

Let $br_2(p_+, p_-)$ be 2's best reply to (p_+, p_-).

$$br_2(p_+, p_-) = \begin{cases} NN, & \text{if } 1-\gamma(p_+, p_-) < \alpha < \gamma(p_+, p_-) \\ BB, & \text{if } \gamma(p_+, p_-) < \alpha < 1-\gamma(p_+, p_-) \\ F, & \text{if } \alpha > \max\{\gamma(p_+, p_-), 1-\gamma(p_+, p_-)\} \end{cases} \tag{9}$$

By Claim 1, 2 in equilibrium mixes either F and BB or F and NN.

Case 1. Consider an equilibrium where 2 mixes F and BB with probability q and $1-q$, respectively, $0 < q < 1$.

In this case, 2 is indifferent between playing F and BB. By (9), $\alpha = 1 - \gamma(p_+, p_-)$. Namely,

$$\alpha = \frac{w_2(1-\frac{1}{2}\beta)gp_+ + w_2(1-g)p_-}{(1-r_2)[g(1-p_+)+(1-g)(1-p_-)] + w_2[gp_+(1-\beta)+(1-g)p_-]} \tag{10}$$

Since $\alpha \in (\frac{1}{2}, 1)$, the equilibrium (p_+^*, p_-^*) satisfies $1-\gamma(p_+^*, p_-^*) \in (\frac{1}{2}, 1)$.

2 obtains the same expected payoff whether she plays F or BB, which is

$$\Pi_2(p_+, p_-) = r_2[g(1-p_+)+(1-g)(1-p_-)] + w_2[gp_+ + (1-g)p_-] \tag{11}$$

Given 2's strategy $(q, 1-q)$, 1's payoffs by playing each pure strategy are

$$\Pi_1^+(NH|q) = q[\alpha w_1 + (1-\alpha)r_1] + (1-q)r_1 \tag{12}$$
$$\Pi_1^+(H|q) = q(1-\alpha') \tag{13}$$
$$\Pi_1^-(NH|q) = q[\alpha w_1 + (1-\alpha)r_1] + (1-q)r_1 \tag{14}$$
$$\Pi_1^-(H|q) = q(1-\alpha) \tag{15}$$

Case (1. i): Suppose 1^+ is indifferent between NH and H. Equating (12) and (13), we have

$$q^* = \frac{r_1}{1-\alpha'-\alpha(w_1-r_1)}$$

and $q^* \in (0,1)$ iff $\alpha < \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$.

Since $\Pi_1^+(NH|q) = \Pi_1^-(NH|q) = \Pi_1^+(H|q) > \Pi_1^-(H|q)$, 1^- will play pure NH. Namely, $p_- = 0$. In equilibrium, $1-\gamma(p_+^*, 0) \in (\frac{1}{2}, 1)$, by (7), $\frac{1}{2} < \dfrac{w_2(1-\frac{1}{2}\beta)gp_+^*}{(1-r_2)[g(1-p_+^*)+1-g] + w_2gp_+^*(1-\beta)} < 1$. Equivalently, $\dfrac{1-r_2}{(1-r_2+w_2)g} < p_+^* < \dfrac{1-r_2}{(1-r_2+\frac{1}{2}\beta w_2)g}$. To guarantee that the last inequality is not void, $\dfrac{1-r_2}{(1-r_2+w_2)g} < 1$, namely, $g > \frac{1-r_2}{1-r_2+w_2}$ must hold.

Substituting $p_-^* = 0$ in (10), we have

$$p_+^* = \frac{\alpha(1-r_2)}{[(1-\alpha')w_2 + (1-r_2)\alpha]g}, \;\; p_-^* = 0 \tag{16}$$

and $0 < p_+^* < 1$ iff $\alpha < \frac{(1-\frac{1}{2}\beta)gw_2}{(1-r_2)(1-g)+(1-\beta)w_2g}$.

We conclude that Case 1.1 is s.t. 1^+mixes H and NH, 1^- plays pure NH and 2 mixes F and BB. The constraints are $\alpha < \max\{\frac{(1-\frac{1}{2}\beta)gw_2}{(1-r_2)(1-g)+(1-\beta)w_2g}, \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}\}$ and $g > \frac{1-r_2}{1-r_2+w_2}$.

By (11) and (16),

$$\Pi_2^* = w_2 \cdot \frac{(1-\alpha')r_2 + \alpha(1-r_2)}{(1-\alpha')w_2 + \alpha(1-r_2)} \tag{17}$$

and $\frac{\partial \Pi_2^*}{\partial \alpha} = w_2 \cdot \frac{(1-r_2)(w_2-r_2)(1-\frac{\beta}{2})}{[\alpha(1-r_2)+(1-\alpha')w_2]^2} > 0$, $\frac{\partial \Pi_2^*}{\partial \beta} = -w_2 \cdot \frac{\alpha(1-r_2)(w_2-r_2)(\alpha-\frac{1}{2})}{[\alpha(1-r_2)+(1-\alpha')w_2]^2} < 0$. 1 of both types has expected payoff

$$\Pi_1^{+*} = \Pi_1^{-*} = \frac{r_1(1-\alpha')}{1-\alpha'-\alpha(w_1-r_1)} \tag{18}$$

and $\frac{\partial \Pi_1^{+*}}{\partial \alpha} = \frac{\partial \Pi_1^{-*}}{\partial \alpha} = \frac{r_1(w_1-r_1)(1-\frac{\beta}{2})}{[1-\alpha'-\alpha(w_1-r_1)]^2} > 0$, $\frac{\partial \Pi_1^{+*}}{\partial \beta} = \frac{\partial \Pi_1^{-*}}{\partial \beta} = -\frac{\alpha r_1(w_1-r_1)(\alpha-\frac{1}{2})}{[1-\alpha'-\alpha(w_1-r_1)]^2} < 0$ That is, all payoffs are decreasing in β and increasing in α.

Case (1. ii): Suppose 1^- is indifferent between NH and H. Equating (14) and (15), we have

$$q^* = \frac{r_1}{1-\alpha-\alpha(w_1-r_1)}$$

Note that $q^* \in (0,1)$ iff $\alpha < \frac{1-r_1}{1-r_1+w_1}$.

Since $\Pi_1^+(H|q) > \Pi_1^-(H|q) = \Pi_1^-(NH|q) = \Pi_1^+(NH|q)$, 1^+ will play pure H, namely, $p_+ = 1$. In this case, $1-\gamma(1,p_-^*) \in (\frac{1}{2},1)$, by (7), $\frac{1}{2} < \frac{w_2(1-\frac{1}{2}\beta)g + w_2(1-g)p_-^*}{(1-r_2)(1-g)(1-p_-^*)+w_2[g(1-\beta)+(1-g)p_-^*]} < 1$. Equivalently, $\frac{(1-r_2)(1-g)-w_2 g}{(1-r_2+w_2)(1-g)} < p_-^* < \frac{(1-r_2)(1-g)-\frac{1}{2}\beta w_2 g}{(1-r_2+w_2)(1-g)}$. To guarantee that the last inequality is not void, $\frac{1}{2}\beta w_2 g < (1-r_2)(1-g)$, namely, $g < \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2}$ must hold. By (10), we have

$$p_+^* = 1, \;\; p_-^* = \frac{(1-r_2)(1-g)\alpha - w_2(1-\alpha')g}{(1-g)[w_2+(1-r_2-w_2)\alpha]}$$

and $0 < p_-^* < 1$ iff $\alpha > \frac{(1-\frac{1}{2}\beta)gw_2}{(1-r_2)(1-g)+(1-\beta)w_2 g}$.

We conclude that the equilibrium of Case 1.2 is s.t. 1^+ plays pure H, 1^- mixes H and NH and 2 mixes F and BB. The constraints are $\frac{(1-\frac{1}{2}\beta)gw_2}{(1-r_2)(1-g)+(1-\beta)w_2 g} < \alpha < \frac{1-r_1}{1-r_1+w_1}$ and $g < \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2}$.

By (11),

$$\Pi_2^* = w_2 \cdot \frac{(w_2-r_2)\beta g(\frac{1}{2}-\alpha) + (1-2\alpha)r_2 + \alpha}{w_2+(1-r_2-w_2)\alpha} \tag{19}$$

and $\frac{\partial \Pi_2^*}{\partial \alpha} = w_2(w_2-r_2) \cdot \frac{(1-r_2)(1-\frac{1}{2}\beta g)-\frac{1}{2}\beta w_2 g}{[w_2+(1-r_2-w_2)\alpha]^2} > w_2(w_2-r_2) \cdot \frac{(1-r_2)(1-g)-\frac{1}{2}\beta w_2 g}{[w_2+(1-r_2-w_2)\alpha]^2} > 0$, $\frac{\partial \Pi_2^*}{\partial \beta} = -\frac{w_2(w_2-r_2)(\alpha-\frac{1}{2})g}{w_2+(1-r_2-w_2)\alpha} < 0$. In this case, 1 of both types has expected payoff

$$\Pi_1^{+*} = \frac{r_1(1-\alpha')}{1-\alpha-\alpha(w_1-r_1)}, \;\; \Pi_1^{-*} = \frac{r_1(1-\alpha)}{1-\alpha-\alpha(w_1-r_1)} \tag{20}$$

and $\frac{\partial \Pi_1^{+*}}{\partial \alpha} = \frac{r_1[(1-\frac{1}{2}\beta)(w_1-r_1)+\frac{1}{2}\beta]}{[1-\alpha-\alpha(w_1-r_1)]^2} > 0$, $\frac{\partial \Pi_1^{-*}}{\partial \alpha} = \frac{r_1(w_1-r_1)}{[1-\alpha-\alpha(w_1-r_1)]^2} > 0$, $\frac{\partial \Pi_1^{+*}}{\partial \beta} = \frac{r_1(\alpha-\frac{1}{2})}{1-\alpha-\alpha(w_1-r_1)} > 0$, and $\frac{\partial \Pi_1^{-*}}{\partial \beta} = 0$. All payoffs are increasing in α. The payoffs of 1^+ and 2 are decreasing in β and the payoff of 1^- does not depend on β.

Case 2. Suppose 2 mixes F and NN with probability q and $1-q$, respectively.

In equilibrium, 2 is indifferent between F and NN. By (9), the equilibrium (p_+, p_-) satisfies $\alpha = \gamma(p_+, p_-)$. That is

$$\alpha = \frac{(1-r_2)[g(1-p_+) + (1-g)(1-p_-)] - \frac{1}{2}\beta g p_+ w_2}{(1-r_2)[g(1-p_+) + (1-g)(1-p_-)] + w_2[g p_+(1-\beta) + (1-g)p_-]} \tag{21}$$

2's payoff is identical whether she plays F or NN, which is

$$\Pi_2(p_+, p_-) = g(1-p_+) + (1-g)(1-p_-) \tag{22}$$

Conditioned on 2's strategy q, 1's payoffs by playing each pure strategy are

$$\Pi_1^+(NH|q) = q[\alpha w_1 + (1-\alpha)r_1] + (1-q)w_1 \tag{23}$$
$$\Pi_1^+(H|q) = q(1-\alpha') + (1-q) \tag{24}$$
$$\Pi_1^-(NH|q) = q[\alpha w_1 + (1-\alpha)r_1] + (1-q)w_1 \tag{25}$$
$$\Pi_1^-(H|q) = q(1-\alpha) + (1-q) \tag{26}$$

Case (2. i): Suppose 1^+ is indifferent between NH and H. Equating (23) and (24), we have

$$q^* = \frac{1-w_1}{\alpha' - (1-\alpha)(w_1-r_1)}$$

and $q^* \in (0,1)$ iff $\alpha > \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$. Since $\Pi_1^-(H|q) < \Pi_1^+(H|q) = \Pi_1^+(NH|q) = \Pi_1^-(NH|q)$, 1^- prefers NH to H, namely, $p_-^* = 0$ and both types of 1 have the same expected payoff

$$\Pi_1^{+*} = \Pi_1^{-*} = \frac{w_1\alpha' - (1-\alpha)(w_1-r_1)}{\alpha' - (1-\alpha)(w_1-r_1)} \tag{27}$$

Since $p_- = 0$, in equilibrium, $\gamma(p_+^*, 0) \in (\frac{1}{2}, 1)$, by (7),

$$\frac{1}{2} < \frac{(1-r_2)(1-gp_+^*) - \frac{1}{2}\beta w_2 g p_+^*}{(1-r_2)(1-gp_+^*) + w_2 g p_+^*(1-\beta)} < 1. \text{ This is equivalent to } p_+^* < \frac{1-r_2}{(1-r_2+w_2)g}.$$

By (21) we have

$$p_+^* = \frac{(1-\alpha)(1-r_2)}{w_2\alpha' g + (1-\alpha)(1-r_2)g}, \quad p_-^* = 0$$

and $0 < p_+^* < 1$ iff $\alpha > 1 - \frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}$. By (22),

$$\Pi_2^* = \frac{w_2\alpha'}{w_2\alpha' + (1-\alpha)(1-r_2)} \tag{28}$$

We have $\frac{\partial \Pi_1^{+*}}{\partial \alpha} = \frac{\partial \Pi_1^{-*}}{\partial \alpha} = \frac{(1-w_1)(w_1-r_1)(1-\frac{1}{2}\beta)}{[\alpha'-(1-\alpha)(w_1-r_1)]^2} > 0$,
$\frac{\partial \Pi_1^{+*}}{\partial \beta} = \frac{\partial \Pi_1^{-*}}{\partial \beta} = -\frac{(1-w_1)(w_1-r_1)(\alpha-\frac{1}{2})(1-\alpha)}{[\alpha'-(1-\alpha)(w_1-r_1)]^2} < 0$, $\frac{\partial \Pi_2^*}{\partial \alpha} = \frac{w_2(1-r_2)(1-\frac{1}{2}\beta)}{[w_2\alpha'+(1-\alpha)(1-r_2)]^2} > 0$ and
$\frac{\partial \Pi_2^*}{\partial \beta} = -\frac{w_2(1-r_2)(1-\alpha)(\alpha-\frac{1}{2})}{[w_2\alpha'+(1-\alpha)(1-r_2)]^2} < 0$.

We conclude that in this case the equilibrium is s.t. 1^+ mixes H and NH, 1^- plays pure NH and 2 mixes F and NN. The constraints are $\alpha > \max\{1 - \frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}, \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}\}$. All the payoffs are increasing in α and decreasing in β.

Case (2. ii): Suppose 1^- is indifferent with NH and H. Then 1^+ prefers H over NH. Equating (25) and (26), we have

$$q^* = \frac{1-w_1}{\alpha - (1-\alpha)(w_1-r_1)}$$

and $q^* \in (0,1)$ iff $\alpha > \frac{1-r_1}{1-r_1+w_1}$.

$$\Pi_1^{+*} = 1 - \frac{(1-w_1)\alpha'}{\alpha-(1-\alpha)(w_1-r_1)}, \ \Pi_1^{-*} = \frac{w_1\alpha - (1-\alpha)(w_1-r_1)}{\alpha-(1-\alpha)(w_1-r_1)} \tag{29}$$

$\frac{\partial \Pi_1^{+*}}{\partial \alpha} = (1-w_1)\frac{(w_1-r_1)(1-\frac{1}{2}\beta)+\frac{1}{2}\beta}{[\alpha-(1-\alpha)(w_1-r_1)]^2} > 0$, $\frac{\partial \Pi_1^{-*}}{\partial \alpha} = \frac{(1-w_1)(w_1-r_1)}{[\alpha-(1-\alpha)(w_1-r_1)]^2} > 0$,
$\frac{\partial \Pi_1^{+*}}{\partial \beta} = \frac{(1-w_1)(1-\alpha)}{\alpha-(1-\alpha)(w_1-r_1)} < 0$ and $\frac{\partial \Pi_1^{-*}}{\partial \beta} = 0$.

As $p_+ = 1$ and $\gamma(1, p_-^*) \in (\frac{1}{2}, 1)$, by (7), $\frac{1}{2} < \frac{(1-r_2)(1-g)(1-p_-^*)-\frac{1}{2}\beta w_2 g}{(1-r_2)(1-g)(1-p_-^*)+w_2[g(1-\beta)+(1-g)p_-^*]} < 1$.
Equivalently, $p_-^* < \frac{(1-r_2)(1-g)-w_2 g}{(1-r_2+w_2)(1-g)}$. To guarantee that $p_-^* \in (0,1)$, $0 \leq g < \frac{1-r_2}{1-r_2+w_2}$ must hold.
By (21) we have

$$p_+^* = 1, \ \ p_-^* = \frac{(1-\alpha)(1-r_2)(1-g) - w_2\alpha' g}{(1-g)[\alpha w_2 + (1-\alpha)(1-r_2)]}$$

and $0 < p_-^* < 1$ iff $\alpha < 1 - \frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}$. By (22),

$$\Pi_2^* = \frac{(1-g)\alpha w_2 + w_2\alpha' g}{\alpha w_2 + (1-\alpha)(1-r_2)} \tag{30}$$

and $\frac{\partial \Pi_2^*}{\partial \alpha} = w_2 \frac{(1-\frac{1}{2}\beta g)(1-r_2)-\frac{1}{2}\beta g w_2}{[\alpha w_2+(1-\alpha)(1-r_2)]^2} > w_2 \frac{(1-g)(1-r_2)-g w_2}{[\alpha w_2+(1-\alpha)(1-r_2)]^2} > 0$, $\frac{\partial \Pi_2^*}{\partial \beta} = -\frac{(\alpha-\frac{1}{2})w_2 g}{\alpha w_2+(1-\alpha)(1-r_2)} < 0$.

We conclude that in this case, the equilibrium is s.t. 1^+ plays H, 1^- mixes H and NH and 2 mixes F and NN. The constraints are $\frac{1-r_1}{1-r_1+w_1} < \alpha < 1 - \frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}$ and $0 \leq g < \frac{1-r_2}{1-r_2+w_2}$. The payoffs are all increasing in α. The payoff of 1^+ and 2 are decreasing in β and the payoff of 1^- is independent of β.

We next characterize the equilibrium of Γ_r. To that end, we denote a typical tuple of parameters in $\mathcal{L}$ by $\ell \equiv (g, \alpha, \beta, r_1, w_1, r_2, w_2)$ and define five subsets (regions) of $\mathcal{L}$. The game $\Gamma_r(\ell)$ is the game Γ_r with the parameter profile ℓ.

$$R_{FB.1} \equiv \{\ell \in \mathcal{L} | g > \frac{1-r_2}{1-r_2+w_2} \text{ and } \alpha < \min\{\frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}, \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}\}\}$$

$$R_{FB.2} \equiv \{\ell \in \mathcal{L} | g < \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2} \text{ and } \frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g} < \alpha < \frac{1-r_1}{1-r_1+w_1}\}$$

$$R_{FN.1} \equiv \{\ell \in \mathcal{L} | \alpha > \max\{1 - \frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}, \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}\}\}$$

$$R_{FN.2} \equiv \{\ell \in \mathcal{L} | g < \frac{1-r_2}{1-r_2+w_2} \text{ and } \frac{1-r_1}{1-r_1+w_1} < \alpha < 1 - \frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}\}$$

$$R_F \equiv \{\ell \in \mathcal{L} | \max\{\frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+w_2 g(1-\beta)}, 1 - \frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+w_2 g(1-\beta)}, \frac{1-r_1}{1-r_1+w_1}\} < \alpha < \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}\}$$

Given the complexity of the conditions defining the five regions, it is not obvious that they form a partition of $\mathcal{L}$. To show that indeed they are disjoint and cover all $\mathcal{L}$, we further partition each region as shown in Table 7.

The inequalities in the top row apply to all five regions. Each inequality in the cells, together with its corresponding top row in the same column, defines a subset of a region. For instance, suppose $g < \frac{1-r_2}{1-r_2+w_2}$ and $1 - w_1 > r_1$. Then there is no $\ell \in R_{FB.1}$. Suppose$g < \frac{1-r_2}{1-r_2+w_2}$, $1 - w_1 > r_1$ and $\alpha < \frac{1-r_1}{1-r_1+w_1}$. Then $\ell \in R_{FB.2}$.

Summing up the results of Proposition 1 and 2, we characterize the equilibrium of $\Gamma_r(\ell)$ in Table 8.

Table 7: Partition of $\mathcal{L}$ into five regions

	$g > \frac{1-r_2}{1-r_2+w_2}$ and $1-w_1 > r_1$	$g < \frac{1-r_2}{1-r_2+w_2}$ and $1-w_1 > r_1$
$\ell \in R_{FB.1}$	$\alpha < \min\{\frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}, \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}\}$	$\emptyset$
$\ell \in R_{FB.2}$	$g < \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2}$ and $\frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g} < \alpha < \frac{1-r_1}{1-r_1+w_1}$	$\alpha < \frac{1-r_1}{1-r_1+w_1}$
$\ell \in R_{FN.1}$	$\alpha > \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$	$\alpha > \max\{1-\frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}, \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}\}$
$\ell \in R_{FN.2}$	$\emptyset$	$\frac{1-r_1}{1-r_1+w_1} < \alpha < 1-\frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}$
$\ell \in R_F$	$\max\{\frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+w_2 g(1-\beta)}, \frac{1-r_1}{1-r_1+w_1}\} < \alpha < \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$	$\max\{1-\frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+w_2 g(1-\beta)}, \frac{1-r_1}{1-r_1+w_1}\} < \alpha < \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$

	$g = \frac{1-r_2}{1-r_2+w_2}$ and $1-w_1 > r_1$	$g < \frac{1-r_2}{1-r_2+w_2}$ and $1-w_1 \le r_1$	$g \ge \frac{1-r_2}{1-r_2+w_2}$ and $1-w_1 \le r_1$
$\ell \in R_{FB.1}$	$\emptyset$	$\emptyset$	$\emptyset$
$\ell \in R_{FB.2}$	$\alpha < \frac{1-r_1}{1-r_1+w_1}$	$\emptyset$	$\emptyset$
$\ell \in R_{FN.1}$	$\alpha > \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$	$\alpha > 1-\frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}$	$\mathcal{L}$
$\ell \in R_{FN.2}$	$\emptyset$	$\alpha < 1-\frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}$	$\emptyset$
$\ell \in R_F$	$\frac{1-r_1}{1-r_1+w_1} < \alpha < \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$	$\emptyset$	$\emptyset$

Table 8: Properties of the Equilibrium of $\Gamma_r(\ell)$

	$g > \frac{1-r_2}{1-r_2+w_2}$ and $\alpha < \min\{\frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}, \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}\}$ $\quad (R_{FB.1})$	$g < \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2}$ and $\frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g} < \alpha < \frac{1-r_1}{1-r_1+w_1}$ $\quad (R_{FB.2})$
2's equilibrium strategy	$Pr_2(F) = q^* = \frac{r_1}{1-\alpha'-\alpha(w_1-r_1)}$, $Pr_2(BB) = 1-q^*$	$Pr_2(F) = q^* = \frac{r_1}{1-\alpha-\alpha(w_1-r_1)}$, $Pr_2(BB) = 1-q^*$
1^+'s equilibrium strategy	$p_+^* = \frac{\alpha(1-r_2)}{g[(1-\alpha')w_2+(1-r_2)\alpha]}$	$p_+^* = 1$ (Pure (H,D))
1^-'s equilibrium strategy	$p_-^* = 0$ (Pure NH)	$p_-^* = \frac{(1-r_2)(1-g)\alpha - w_2(1-\alpha')g}{(1-g)[w_2+(1-r_2-w_2)\alpha]}$
2's payoff	$\Pi_2^* = w_2 \cdot \frac{(1-\alpha')r_2+\alpha(1-r_2)}{(1-\alpha')w_2+\alpha(1-r_2)}$	$\Pi_2^* = w_2 \cdot \frac{\alpha-(2\alpha-1)r_2-(w_2-r_2)g(\alpha-\alpha')}{(1-\alpha)w_2+\alpha(1-r_2)}$
1^+'s payoff	$\Pi_1^{+*} = \frac{r_1(1-\alpha')}{1-\alpha'-\alpha(w_1-r_1)}$	$\Pi_1^{+*} = \frac{r_1(1-\alpha')}{1-\alpha-\alpha(w_1-r_1)}$
1^-'s payoff	$\Pi_1^{-*} = \frac{r_1(1-\alpha')}{1-\alpha'-\alpha(w_1-r_1)}$	$\Pi_1^{-*} = \frac{r_1(1-\alpha)}{1-\alpha-\alpha(w_1-r_1)}$
if α increases	$\Pi_2^* \uparrow$, $\Pi_1^{+*} = \Pi_1^{-*} \uparrow$	$\Pi_2^* \uparrow$, $\Pi_1^{+*} \uparrow$, $\Pi_1^{-*} \uparrow$
if β increases	$\Pi_2^* \downarrow$, $\Pi_1^{+} = \Pi_1^{-*} \downarrow$	$\Pi_2^* \downarrow$, $\Pi_1^{+*} \uparrow$, $\Pi_1^{-*}-$
	$\alpha > \max\{1 - \frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}, \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}\}$ $\quad (R_{FN.1})$	$g < \frac{1-r_2}{1-r_2+w_2}$ and $\frac{1-r_1}{1-r_1+w_1} < \alpha < 1 - \frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+(1-\beta)w_2 g}$ $\quad (R_{FN.2})$
2's equilibrium strategy	$Pr_2(F) = q^* = \frac{1-w_1}{\alpha'-(1-\alpha)(w_1-r_1)}$, $Pr_2(NN) = 1-q^*$	$Pr_2(F) = q^* = \frac{1-w_1}{\alpha-(1-\alpha)(w_1-r_1)}$, $Pr_2(NN) = 1-q^*$
1^+'s equilibrium strategy	$p_+^* = \frac{(1-\alpha)(1-r_2)}{g[w_2\alpha'+(1-r_2)(1-\alpha)]}$	$p_+^* = 1$ (Pure (H,D))
1^-'s equilibrium strategy	$p_-^* = 0$ (Pure NH)	$p_-^* = \frac{(1-\alpha)(1-r_2)(1-g)-w_2\alpha' g}{(1-g)[\alpha w_2+(1-\alpha)(1-r_2)]}$
2's payoff	$\Pi_2^* = \frac{w_2\alpha'}{w_2\alpha'+(1-\alpha)(1-r_2)}$	$\Pi_2^* = w_2 \cdot \frac{(1-g)\alpha+\alpha' g}{\alpha w_2+(1-\alpha)(1-r_2)}$
1^+'s payoff	$\Pi_1^{+*} = \frac{w_1\alpha'-(1-\alpha)(w_1-r_1)}{\alpha'-(1-\alpha)(w_1-r_1)}$	$\Pi_1^{+*} = 1 - \frac{(1-w_1)\alpha'}{\alpha-(1-\alpha)(w_1-r_1)}$
1^-'s payoff	$\Pi_1^{-*} = \frac{w_1\alpha'-(1-\alpha)(w_1-r_1)}{\alpha'-(1-\alpha)(w_1-r_1)}$	$\Pi_1^{-*} = \frac{w_1\alpha-(1-\alpha)(w_1-r_1)}{\alpha-(1-\alpha)(w_1-r_1)}$
if α increases	$\Pi_2^* \uparrow$, $\Pi_1^{+*} = \Pi_1^{-*} \uparrow$	$\Pi_2^* \uparrow$, $\Pi_1^{+*} \uparrow$, $\Pi_1^{-*} \uparrow$
if β increases	$\Pi_2^* \downarrow$, $\Pi_1^{+*} = \Pi_1^{-*} \downarrow$	$\Pi_2^* \downarrow$, $\Pi_1^{+*} \uparrow$, $\Pi_1^{-*}-$
	$\max\{\frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+w_2 g(1-\beta)}, 1 - \frac{(1-\frac{1}{2}\beta)w_2 g}{(1-r_2)(1-g)+w_2 g(1-\beta)}, \frac{1-r_1}{1-r_1+w_1}\} < \alpha < \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$	
2's equilibrium strategy	$Pr_2(F) = 1$ (Pure F)	
1^+'s equilibrium strategy	$p_+^* = 1$ (Pure (H,D))	
1^-'s equilibrium strategy	$p_-^* = 0$ (Pure NH)	
2's payoff	$\Pi_2^* = (1-g)[\alpha + (1-\alpha)r_2] + g\alpha' w_2$	
1^+'s payoff	$\Pi_1^{+*} = 1-\alpha'$	
1^-'s payoff	$\Pi_1^{-*} = \alpha w_1 + (1-\alpha)r_1$	
if α increases	$\Pi_2^* \uparrow$, $\Pi_1^{+*} \downarrow$, $\Pi_1^{-*} \uparrow$	
if β increases	$\Pi_2^* \downarrow$, $\Pi_1^{+*} \uparrow$, $\Pi_1^{-*}-$	

□

Proof of Proposition 3. (I.) Suppose g=1. That is, it is common knowledge that 1 is of type 1^+. Then

$$\gamma(p_+) = \frac{(1-r_2)(1-p_+) - \frac{1}{2}\beta p_+ w_2}{(1-r_2)(1-p_+) + (1-\beta)p_+ w_2}$$

2's best reply is

$$br_2(p_+) = \begin{cases} NN, & \text{if } 1-\gamma(p_+) < \alpha < \gamma(p_+) \\ BB, & \text{if } \gamma(p_+) < \alpha < 1-\gamma(p_+) \\ F, & \text{if } \alpha > \max[\gamma(p_+), 1-\gamma(p_+)] \end{cases}$$

Consider an equilibrium where 2 mixes F and BB with probability q and $1-q$, respectively, $0 < q < 1$.

In this case, 2 is indifferent between playing F and BB. By (9), $\alpha = 1-\gamma(p_+)$ and

$$\alpha = \frac{w_2(1-\frac{1}{2}\beta)p_+}{(1-r_2)(1-p_+) + w_2 p_+(1-\beta)}$$

equivalently, $p_+^* = \frac{\alpha(1-r_2)}{(1-\alpha')w_2 + \alpha(1-r_2)}$.

2 obtains the same expected payoff whether she plays F or BB, which is

$$\Pi_2^* = r_2(1-p_+) + w_2 p_+ = w_2 \cdot \frac{(1-\alpha')r_2 + \alpha(1-r_2)}{(1-\alpha')w_2 + \alpha(1-r_2)}$$

1 is indifferent with NH and H.

$$\Pi_1^+(NH|q) = q[\alpha w_1 + (1-\alpha)r_1] + (1-q)r_1 = \Pi_1^+(H|q) = q(1-\alpha')$$

The solution in q is $q^* = \frac{r_1}{1-\alpha'-\alpha(w_1-r_1)}$ and $Pr_2(BB) = 1-q^*$. Note that $q^* \in (0,1)$ iff $\alpha < \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$.

1's expected payoff is $\Pi_1^* = \frac{r_1(1-\alpha')}{1-\alpha'-\alpha(w_1-r_1)}$.
Consider next an equilibrium where 2 mixes F and NN with probability q and $1-q$, respectively.

Since 2 is indifferent between F and NN, by (9), $\alpha = \gamma(p_+)$. That is

$$\alpha = \frac{(1-r_2)(1-p_+) - \frac{1}{2}\beta p_+ w_2}{(1-r_2)(1-p_+) + w_2 p_+(1-\beta)}$$

The solution in p_+ is $p_+^* = \frac{(1-\alpha)(1-r_2)}{\alpha' w_2 + (1-\alpha)(1-r_2)}$.

In equilibrium, 2 obtains the same payoff whether she chooses F or NN. Hence $\Pi_2^* = 1 - p_+^* = \frac{\alpha' w_2}{\alpha' w_2 + (1-\alpha)(1-r_2)}$.

Since 1 is indifferent between H and NH, $\Pi_1^+(NH) = q[\alpha w_1 + (1-\alpha) r_1] + (1-q) w_1 = \Pi_1^+(H) = q(1-\alpha') + (1-q)$. Solving for q, we have $q^* = \frac{1-w_1}{\alpha' - (1-\alpha)(w_1 - r_1)}$. Note that $q^* \in (0,1)$ iff $\alpha > \frac{1-r_1-\frac{1}{2}\beta}{1-r_1+w_1-\beta}$. 1's expected payoff is $\Pi_1^* = 1 - \frac{(1-w_1)\alpha'}{\alpha' - (1-\alpha)(w_1-r_1)}$.

(II.) Suppose $g = 0$. It is equivalent to the case where $g = 1$ and $\beta = 0$. Intuitively, not possessing a disruptive technology ($g = 0$) is equivalent to the case where 1 possesses a disruptive technology ($g = 1$) that is obsolete ($\beta = 0$).

(III.) Comparing the complete information case ($g = 0$ or 1) with the asymmetric information case ($0 < g < 1$). It is easy to verify (using Table 5) that the equilibrium is continuous in g also at $g = 0$ and $g = 1$. □