What the Enemy Knows
Common Knowledge and the Rationality of War

Research note

Thomas Chadefaux

Trinity College Dublin
Department of Political Science
College Green 2, 4.02
Dublin 2, Ireland
E-mail: chadefat@tcd.ie
http://www.thomaschadefaux.com

October 13, 2015

Abstract

Information has played a central role in understanding why international negotiations may break down into costly conflict. Barring indivisibilities or commitment problems, the literature finds that war can only occur between rational unitary actors because of private information about fundamentals such as capabilities, resolve, or costs. I show here, however, that negotiations may fail despite complete information about these fundamentals. All that is needed is for A to not know whether B knows—uncertainty about uncertainty. To ensure peace, then, states need not only know each other's attributes, but also the other's knowledge thereof, and potentially his knowledge of her own knowledge, and so on. Existing models, however, focus on first-order uncertainty and assume common knowledge of information partitions—an unlikely assumption, as states rarely know how much the other knows. This requirement of higher-order complete information illustrates the importance of explicitly incorporating information structures in bargaining models of conflict.

Keywords: Bargaining, conflict, war, common knowledge, uncertainty, information

Introduction

The onset of costly interstate wars despite the existence of more efficient negotiated solutions is a longstanding puzzle in international conflict research. Much of the literature points to the role of private information and incentives to misrepresent as causes for these bargaining failures. The argument is typically that one or both states misestimate resolve, preferences, costs, or the distribution of capabilities, and this asymmetric information can lead both to expect a positive utility for war.

Yet the asymmetric information included in crisis bargaining models is typically limited to payoff-relevant events—fundamentals such as capabilities, costs or preferences. What players do or do not know, on the other hand, is almost always assumed to be common knowledge.[1] There is, in other words, no uncertainty about the presence of uncertainty. States may not know what the other has, but they know whether she knows. In most real-world situations, however, states rarely know how much information the other has. They need to guess whether their opponent's spies succeeded in acquiring secret information (and hence whether their counterintelligence efforts have been successful), and whether the signals sent were interpreted correctly. What the other knows, in other words, is itself private and potentially strategic

[1] We say that a proposition is *common knowledge* if every player knows it, every player knows that every player knows it, and so on ad infinitum (Aumann 1976). A player's *private information* is any information that he has that is not common knowledge among all players in the game.

information.

We show here that bargaining may in fact break down into war simply because of uncertainty about uncertainty itself, even if both states completely know each other's capabilities, costs and resolve. This goes against the widespread idea that complete information alone is sufficient to ensure peace. The intuition for this result is simple: suppose that A thinks B might be mistaken about the distribution of power (even if she is not). For example, A might think that B thinks A has developed nuclear capabilities—i.e., that the balance of power is less favourable to B than it actually is. If it is sufficiently probable that B is mistaken, then A has an incentive to offer her a small share of the pie. But because there is a chance that B actually knows the truth, B might in fact prefer war to that offer.

If war is caused by incomplete information about fundamental attributes such as capabilities, resolve, or costs, then transparency or credible signals should solve the problem of inefficient bargaining outcomes. But if complete information about the distribution of information is also necessary, then the conditions for the absence of conflict are more stringent than may have been thought. In particular, it implies that states must not only gather and convey first-order information—about their opponents's capabilities and resolve—but also second- or higher-order information: information about what the opponent knows about her, and possibly what he knows she knows about her, and so on.

In sum, this research note shows the importance of higher-order uncertainty for the onset of war. War may occur between two rational and unitary actors despite complete information about capabilities or resolve if players are unsure of whether the other knows the distribution of power and resolve. The paper proceeds in three steps. First, the role of information and common knowledge in the bargaining literature is discussed. Second, we present a simple model of complete information about fundamentals, without problems of indivisibility or commitment, in which war nonetheless occurs with positive probability in all equilibria due to uncertainty about information partitions. Having established the importance of higher-order information, we then address in a third section the challenges associated with reaching common knowledge.

Uncertain Uncertainty

The role of information in interstate relationships has been at the core of the research on conflict over the past 30 years. The central explanation for the onset of war between rational actors is the idea that at least one player has incomplete information about some of his opponents' attributes such as their capabilities, resolve or costs for war (Jervis 1976, Blainey 1988, Fearon 1995, Powell 1996, Van Evera 1999, Reiter 2003).[2] Private information and

[2] Of course, not all of the bargaining literature is related to incomplete information—see for example rationalist explanations for war related to commitment problems (Powell 2004*b*, Chadefaux 2011, Chadefaux 2012) or indivisibilities (Toft 2006). Gartzke (1999), however, argues that the three rationalist explanations presented in Fearon (1995)—private

incentives to misrepresent it, in turn, lead to misperceptions or miscalculations about the distribution of power or resolve. Negotiators may thus be optimistic about the expected outcome of a war, with the result that both believe they have a higher expected value for war than for peace (Fey & Ramsay 2007, Slantchev & Tarar 2011).[3]

The type of uncertainty that is typically modeled, however, relates only to payoffs. The games include some uncertainty space corresponding to payoff-relevant attributes, together with the information players have about these attributes. Blainey's often-cited argument, for example, is that "wars usually begin when two nations disagree on their relative strength" (Blainey 1988, p. 246). Similarly Fearon (1995, p. 18) focuses on "disagreements about relative power and uncertainty about a potential opponent's willingness to fight."[4] In these models, players may be unsure of each other's attributes such as the distribution of capabilities or interest, or the cost of war, but

information, commitment problems, and indivisibilities—are all instances of inefficiencies caused by one form or another of incomplete information.

[3]Asymmetric information may also lead wars to last longer than they would otherwise (Wittman 1979, Wagner 2000, Filson & Werner 2002, Slantchev 2003, Powell 2004*a*, Smith & Stam 2004).

[4]Further, Fearon also writes: "state leaders might have private information about militarily relevant factors-military capabilities, strategy, and tactics; the population's willingness to prosecute a long war; or third-state intentions. If a state has superior (and so private) information about any such factor, then its estimate of the probable course of battle may differ from that of an adversary."

they know whether each one is informed of them.[5] In other words, the type of uncertainty that is discussed is of the form "A does not know X, and B knows that A does not know X." The fact that information is asymmetric is common knowledge, so states only face what may be called first-order uncertainty—uncertainty about fundamentals—but not about the information partition itself.[6]

In the real world, however, states rarely know for sure how much the other knows. Intelligence failures and the challenges associated with credibly conveying one's private information imply that states must almost always

[5]Other contributions in this field adopt the same approach: Morrow (1989, p. 942) models a situation in which "each side is uncertain about the other's resolve"; Powell (1996, p. 261) assumes that "each player has private information about its cost of imposing a settlement."; in Fearon (1997, p. 70), "the defender is privately informed of its value for the international interest and then chooses a signal m [...]". In Schultz (1998, p. 834), "each state is uncertain about the other's expected costs from war." In Gartzke (1999, p. 578), "the other country [...] has private information about its costs [of fighting]." Wagner (2000, p. 472) writes that "in the case of war what prevents immediate agreement is not private infozrmation about the personal preferences of political leaders, but conflicting expectations about the relative performance of military forces and/or about the behavior of other actors such as potential allies or political actors within states." Slantchev (2003, p. 624) assumes that "player 1 is uncertain about the distribution of power," and Smith & Stam (2004, p. 787) that "although actors have different beliefs, what these beliefs are is common knowledge to all actors. That is, although nation A has different beliefs than nation B, nation A knows what nation B believes and vice versa."

[6]A recent notable exception is Acharya & Grillo (2015), but their focus is on the absence of common knowledge of *rationality* rather than on higher-order information.

guess what the other has been able to learn or infer. For example, has the adversary discovered my secret nuclear program (or the absence thereof)? Are enemy spies leaking information? Is my phone tapped? And what is made of that information? Did they believe it? And did they think that I thought they believed it?

Can bargaining remain efficient if everything is known except whether the other knows? In other words, can war be an equilibrium outcome in a game of complete information about payoff-relevant events, but not about the information partition—i.e., what the other is informed of?[7] We know from economics that uncertainty about what others know matters.[8] Yet of interest in the literature has been the opposite scenario: one in which information is *incomplete* and trade never happens because *common knowledge* allows the actors to infer from an offer that the other must know something they do not, or else they would not want to trade (Milgrom & Stokey 1982, Fey & Ramsay 2007). Through a potentially infinite regress, they infer that the expected utility of the deal cannot be larger than the one of no trade, and hence trade never occurs.[9] This led to an analogous result in the conflict

[7]Information is complete when each agent knows the other agent's utility function and rules of the game. However, this definition does not include the mutual awareness of the players, which is called 'common knowledge'.

[8]Rubinstein (1989), for example, demonstrates that even Nash equilibria which survive the usual refinements may not be robust to a small lack of common knowledge.

[9]Milgrom & Stokey (1982)'s 'no-trade' result considers a situation in which traders have asymmetric information but share common knowledge that a trade is mutually acceptable. In that case, no trade will take place between traders since "the mere willingness of the

literature. Fey & Ramsay (2007) argue that "the fundamental reason that mutual optimism cannot lead to war is that if both sides are willing to fight, each should infer that they have either underestimated the strength of the opponent or overestimated their own strength. In either case, these inferences lead to a peaceful settlement of the dispute" (p. 738).[10] The point that is being made here, however, is the converse. It is that even with complete information but without common knowledge, war may occur in equilibrium. We now show this with a simple model.

Model

Two states, A and B, negotiate over the partition of a territory of size normalized to one. There are two types of A: a 'strong' A, denoted A^s, would win a conflict against B with probability p^s, whereas a 'weak' A (A^w) would win with probability p^w. A's type is determined by an initial move by nature.

other traders to accept their parts of the bet is evidence to at least one trader that his own part is unfavourable. Hence no trade can be found that is acceptable to all traders. This no-trade result depends crucially on the assumption that it is common knowledge when a trade is carried out that it is feasible and that it is mutually acceptable to all of the participants" (p. 9). See also the related work of Sebenius & Geanakoplos (1983) and Morris (1994).

[10] Slantchev & Tarar (2011) argue, however, that conflict differs from trade in a fundamental way: whereas both parties need to agree to a trade (the equivalent of conflict for our purposes), war on the other hand requires the assent of only one of the two parties. Hence the no-trade theorem does not apply.

Both players know the distribution of power, but A is unsure whether B knows it. To represent this uncertainty, there must therefore be states of the world reached with positive probability in which B knows the distribution of power, and others in which B does not. We represent these by a move by nature, such that instead of having two possible types corresponding to two distributions of power, nature has four possible moves, corresponding to the combination of A's two possible types (w and s) and B's two possible states of knowledge (figure).[11] In short, A observes her own type, but does not know whether B observes A's type. We use parentheses to denote a player's information partition. For example, (x, y) denotes an information partition in which a player knows that she is at either x or y, but not which of the two. With probability p_1, for example, the state is w_1, in which case B observes (w_1) and knows A's type. In this state, however, A only observes (w_1, w_2), and hence assigns positive probability to state w_2 being the true state—a state in which B would not know A's type.

Following Nature's move, we assume for simplicity a take-it or leave-it bargaining protocol in which A makes an offer $x \in [0, 1]$, where x denotes A's proposed share of the territory (and hence $1 - x$ denotes B's share). B observes the offer, which she either accepts or rejects. If B accepts, then

[11] We follow Harsanyi (2004) and model the absence of common knowledge using a game of incomplete information. To be clear, however, we are not introducing incomplete information about any payoff-relevant attributes, but rather incomplete information about incomplete information itself, as a way to model the absence of common knowledge. See also Fudenberg & Tirole (1991, 556–7).

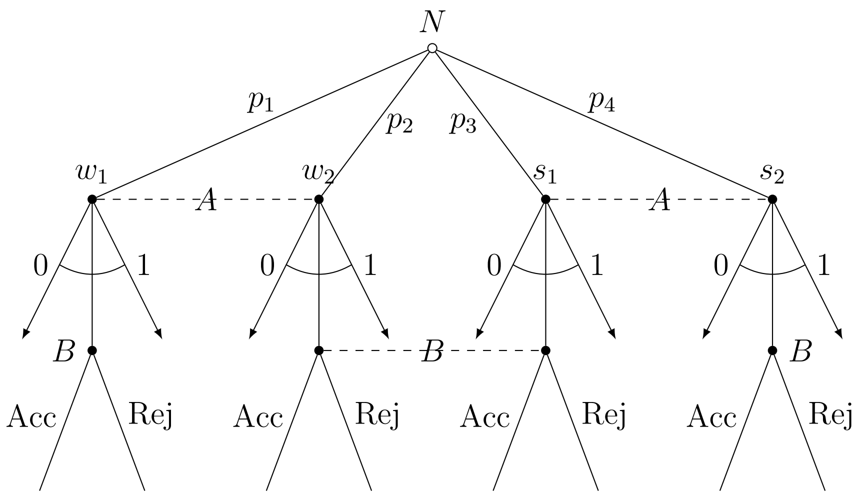

Figure 1: Common Knowledge Game: A does not know whether B knows A's type.

players receive their respective share of the pie and the game ends. If B rejects, however, then war follows, in which case A wins the entire territory with probability p^i (and hence B wins with probability $1 - p^i$), where $i \in \{w, s\}$ denotes A's type, and both players incur cost $c \in (0, 1)$. We assume for simplicity of exposition that both players are risk-neutral (i.e., $u_i(x) = x$).

The intuition for the game's equilibrium is simple. Suppose that A observes (w_1, w_2). This means that A is 'weak', but also that A is not sure whether or not B knows that she is weak, because A assigns positive probability to w_2 being the actual state. Indeed, if w_2 was the true state, B would observe (w_2, s_1), and hence would assign positive probability to the true state of the world being s_1, i.e., to the possibility that A is 'strong'. In other words, upon observing (w_1, w_2), A assigns positive probability to B thinking she is strong. But in that case, i.e., if B thinks A might be strong, then B might be willing to accept a distribution of the pie that reflects this. In turn, this means that A could make an offer x corresponding to a strong type, which B would accept given her beliefs. In equilibrium, therefore, A makes an offer which grants her a large portion of the pie, with the hope that B does not actually observe the true distribution of power. Because B may, however, observe it (since with probability p_1 the state is w_1), war occurs with positive probability. For certain combinations of the parameters, we even show that war occurs with positive probability in all perfect bayesian equilibria of the game.

Whether this can be an equilibrium strategy depends, of course, on the probability of w_1 being the true state of the world given that A observes (w_1, w_2), as well as on s_1 being the true state of the world when B observes (w_2, s_1). In particular, it is crucial that w_1—the probability that B knows given that A observes (w_1, w_2)—be sufficiently unlikely, so that A would want to take the risk to bluff upon observing (w_1, w_2) (or else B would reject the offer and fight), and that s_1 be sufficiently probable for B to accept upon observing (w_2, s_1) (or else she would rather take the risk to reject). We now present the logic of the equilibrium in more detail.

Separating Equilibrium. Consider first the possibility of an equilibrium in which each type of A makes a different offer. Without loss of generality, assume that A offers x^w upon observing (w_1, w_2) but x^s upon observing (s_1, s_2). In this case B always infers from A's offer which type she is facing, and war never occurs. But does such an equilibrium exist? A weak A may be tempted to deviate and offer x^s, which B, given her posterior beliefs, will accept upon observing either (w_2, s_1) or (s_2). The risk that A runs, of course, is that B actually observes (w_1) and hence knows that A is bluffing, in which case war occurs. But if the temptation is sufficiently large and the probability that B discovers the truth (i.e., p_1) sufficiently low, then A will be willing to take that risk and bluff. In that case, A deviates from the equilibrium path, and this can therefore not form the basis of a perfect bayesian equilibrium. There is hence no separating equilibrium (details of the derivation and proofs in the appendix).

Lemma 1. *Assume $p^s - p^w > 2c/q_2^A$, where $q_2^A = p_2/(p_1 + p_2)$. Then there is no peaceful separating perfect bayesian equilibrium.*[12]

This is intuitive. q_2^A is A's posterior probability of being at w_2 upon observing (w_1, w_2). A large q_2^A means a low probability of being caught deviating, and hence that A will be more willing to deviate.

Pooling Equilibrium. Consider now a pooling equilibrium in which both types of A offer x^*. Clearly, x^* must be large enough to satisfy a strong A, as otherwise A would prefer fighting to that agreement. But if x^* is too large, then B will reject it upon observing (w_2, s_1), leading to war—an outcome that neither type of A would like.

[12]A Perfect Bayesian Equilibrium is a pair (s, b) of strategy profile s and beliefs b such that s is sequentially rational given beliefs b, and b is consistent with s. See Fudenberg & Tirole (1991, p. 215) for a more formal definition.

Lemma 2. *Assume* $\frac{2c}{q_2^B} > p^s - p^w > \frac{2c}{q_2^A}$*, where* $q_2^A = \frac{p_2}{p_1+p_2}$ *and* $q_2^B = \frac{p_2}{p_2+p_3}$.[13] *Then there is exists a pooling perfect bayesian equilibrium in which all types of* A *offer* $x^* = p^s - c$*, which* B *accepts unless she observes* (w_1)*, in which case she rejects and war ensues.*

The main proposition then follows immediately: for certain combinations of p^i and nature's moves, there are no peaceful equilibria, and war occurs with positive probability.

Proposition 1. *Assume* $\frac{2c}{q_2^B} > p^s - p^w > \frac{2c}{q_2^A}$*. Then in every pure strategy Perfect Bayesian Nash equilibrium, war occurs with probability* p_1*.*

This implies that even if both A and B know each other's attributes (i.e., w_1 is the true state of the world), war occurs in equilibrium.[14]

Reaching Common Knowledge

The simple model above shows that common knowledge about the distribution of information itself can be a necessary condition for the existence of peaceful equilibria. To avoid war, then, countries need to bridge their perceptions not only of each other's capabilities and resolve, but also of what each knows. Narrowing this gap, however, may prove even more difficult than in the case of first-order asymmetric information. First, the need for

[13]Note that this implicitly assumes that $p_1 \leq p_3$.

[14]There also exist an infinity of semi-separating equilibria, each of which also involves a positive probability of war.

higher order information means that states not only need to send a message, but also to ensure that that message was well received and understood. A message—a 'confirmation'—is therefore also needed, itself subject to further miscommunication and misperception. Second, states have an incentive to misrepresent what they know. Finally, we discuss ways by which states may overcome some of these difficulties and create mechanisms and institutions to create common knowledge.

Learning what the other knows

States acquire information about each other by two main means. The first is to collect it themselves through intelligence services. The second is to rely on what the other conveys in the form of messages and signals. Both intelligence and signals can fail, however. As a result, countries must always wonder: was my signal misunderstood; did the other respond in that way because he does not know or because he knows and still chooses to make that offer? And what has he discovered about myself?

Intelligence. Intelligence involves the use of sensors, spies, or the decryption of messages to collect bits of information that may then be assembled into a coherent whole. Most historical evidence on intelligence is limited to first-order intelligence, which consists in the collection of information about other actors' forces, strategy, or motivation. Yarhi-Milo (2013), for example, studies how leaders assess each other's intentions; May (2014) examines the shortcomings of intelligence before the world wars; Kahn (2000) and Ferris

(2007) focus on intelligence during war. Yet none of them examine higher-order intelligence, which consists in learning what the other knows or believes about you—an area where spies and diplomats play an important role.

Intelligence, however, often fails. Human intelligence relies on interrogations and conversations with those who have information, but may fail due to manipulation, false information being fed to the agent, or the simple inability to penetrate the centre of power.[15] Other sources of failure include cognitive biases or organisational problems (Shore 2005, Garthoff 2015). Signal intelligence and imagery intelligence, which consists of information collected using radars, sonars, or satellites, is limited to what is visible and tangible. Foliage alone may be sufficient to thwart the efforts of a drone or a satellite. Finally, open source intelligence relies on the analysis of journals, radio and television, but is doubly limited: first by the information available to the other country's media, and second by possible manipulation.

These limitations of intelligence gathering of course also apply to the opponent. As a result, it is difficult to know just how much the adversary knows. Has he discovered your secret nuclear program, or the absence thereof? Are

[15]Ferris (2007), for example, notes that "more typically, intelligence consists of news provided five months after the event to His Majesty's Government by British codebreakers of a report from the French ambassador in Bucharest, which he received though the intermediary of a Greek journalist, of the views of Hitler's aims offered by a drunken Japanese charge d'affaires in Sofia. Intelligence services usually provide masses of material, often utterly irrelevant, of unknown accuracy or on a tangent of relevance, drawn from the hearsay of third-hand sources." (p. 102).

spies providing information to the opponent? What is made of that information? And is it believed by the other? Intelligence may also be manipulated to feed false information to the enemy. But did he believe that information? And did he think that I thought he believed it? The mere possibility of intelligence failures, in short, is sufficient to create uncertainty about what the other knows.

Signaling. States can also volunteer information by sending various forms of signals. A large literature in international relations has focused on this strategic exchange of information (Schelling 1980, Fearon 1997, Kydd 2005, Trager 2010), including the use of audience costs (Fearon 1994, but see also Snyder & Borghard (2011)), or even cheap talk (Trager 2010, Crawford & Sobel 1982, Farrell & Gibbons 1989, Sartori 2002). These messages aim to convey information about a state's capabilities and resolve, but also their beliefs about those of the other. Yet signals also often fail. They may fail to be detected or interpreted as informative signals (Jervis 2002, Mercer 2010) and may be affected by a variety of biases (Holsti 1962, Yarhi-Milo 2013, Jervis 2015).

These well-known challenges associated with conveying information are amplified when dealing with common knowledge, because of the need for higher-order levels of information. A signal need not only be sent, but confirmation of its receipt and adequate understanding must also be conveyed back to the sender. The recipient might even have confirmed receipt, but not know whether the sender has received that receipt. Suppose for example that A sends a credible message to B, but that with a small but positive probability the message will be garbled or misinterpreted. Even supposing that the message arrives properly, B still needs to make it clear to A that she received the message. But again, her confirmation and the interpretation might be garbled. So by now A knows, B knows, and A knows that B knows, but B does not know that A knows that B knows. So A needs to send a confirmation, and so on. This problem, known as the coordinated attack problem, illustrates one of the difficulties associated with reaching common knowledge.[16]

To summarize, the transmission of information often fails because of misinterpretation of signals or intelligence failures. The very fallibility of these systems, or even the possibility that they might fail, means that states can hardly ever be certain of what their opponent knows, and common knowledge of information partitions is therefore unlikely. In addition, because the perception of what states know matters, states might not be willing to convey their private information about what they know, even if they could. Indeed, states are likely to be strategic with respect to the information they release information about what they know of the adversary. Just as with first-order incomplete information, states have an incentive to misrepresent their higher-order private information.

[16] See Rubinstein (1989) or a discussion of this problem.

Hiding what you know

Evidence of the strategic importance of higher-order information—information about what I know as opposed to what I have—is reflected in the great care with which states seek to hide what they know or release that information strategically. In the United States, for example, the Presidential Daily Brief is a daily summary of intelligence from various agencies. It is intended to provide the President with key intelligence pertaining to relevant international situations. The daily brief and other information communicated to the President is highly classified and never released to the public (Milbank 2002). The goal is not simply to hide valuable secrets about, say, the geographical position of valuable American assets, but also to hide what the president knows about other countries. What the president knows, or even knew, is highly sensitive and may be used to inform future policy decisions, and hence should not be revealed.

Similarly, at the beginning of the Cuban missile crisis, Kennedy did not want the Soviet Union to know what he knew about the missiles. When he met with Soviet Foreign Minister Andrei Gromyko on October 18th, he was assured that all Soviet assistance to Havana was only "pursued solely for the purpose of contributing to the defense capabilities of Cuba."[17] Kennedy did not then inform him of his knowledge. He continued to avoid any disruption to his public schedule to avoid suspicions that he might know. Later, once the Navy began to enforce the quarantine, however, communications began

[17] John F. Kennedy. *Cuban missile crisis address to the Nation.* October 22, 1962.

to be purposefully sent uncoded. The point then was to ensure that the Soviets would not misperceive their intentions, as well as to ensure American knowledge of their knowledge.

Private information about the distribution of information and the strategic use thereof is also relevant in times of war. When the Enigma code was broken in January 1940 and signals from German communications intercepted, few in the British government and intelligence network were allowed to know. To most, the story was that the information came from an MI6 spy codenamed Boniface. The American allies were not even informed of this breakthrough in intelligence, for fear that the information would leak. It was crucial that Germany not know what the British intelligence services knew. This allowed the British to sink U-boats and avoid attacks over the long run, without raising Germany's suspicion. Had this information leaked, Germany would have changed its encryption system and the UK would have lost its edge.

Finally, second- or higher-order information may also be necessary to infer first-order information. Suppose for example that A finds out that B knows that A is weak. Combined with the fact that B did not attack, A may infer that B is actually weak himself. Clearly, B will therefore want to avoid revealing that he knows that A is weak.

Creating Common Knowledge

There are, however, ways by which states can convey information about attributes such as their military capabilities so that they become common knowledge. They may for instance choose to reveal or display their capabilities publicly. Displays of artillery on national commemoration days, for example, or heavily publicized tests of new weapons thus serve two purposes: signal strength, but also ensure these capabilities are common knowledge. Joint military manoeuvres may also serve the same purpose. By granting the other a more intimate access to her resources, military capabilities become common knowledge.[18]

Similarly, states may want to publicly convey their knowledge of the other's capabilities. They may, for example, mention the precise location of a nuclear site to convey their information in such a way that it becomes common knowledge. In the Cuban missile crisis, for example, satellite photos were used to show the US' knowledge of the hidden ballistic missiles on Cuban soil.

Far more difficult, however, is to generate common knowledge of the *absence* of certain capabilities. Conveying positive knowledge—for example using public displays of force—is far easier than conveying negative knowledge—showing that certain capabilities do not exist. Saddam Hussein thus had a difficult time proving the absence of a program of weapons of mass destruction, whereas the Bush administration could easily pretend

[18]For more on the creation of common knowledge, see also Morrow (2014).

she believed Iraq had one. Similarly, while public displays of force can easily create common knowledge that I have *at least* x units, showing that I have *no more* than z units is far more difficult. Public announcements are only helpful to reveal positive things that only a knowledgeable party would know.

Conclusion

The role of information in bargaining outcomes has been a central theme in existing research on interstate conflict. However, models incorporating asymmetric information have typically limited their attention to private information about fundamentals—first-order incomplete information—while assuming common knowledge of information partitions. Yet, assuming that states know whether their opponent knows the distribution of power or costs is implausible. The very uncertainty that is inherent to information collection—whether it be through intelligence or signaling—implies that states can rarely be certain of what their adversary managed to infer about them. States, in other words, are almost always uncertain of each other's uncertainty.

The importance of relaxing the assumption of common knowledge about information itself was illustrated using a simple bargaining model in which all equilibria included a positive probability of war. Because A thought that B might think she was strong, A had an incentive to bluff and demand large concessions from B. With probability p_1, however, B was well aware of A's weakness, and hence rejected the offer in favour of war.

The importance of higher-order uncertainty also implies that states might behave strategically with regards to how they release or hide information about what they know of the enemy (Prunckun 2012, ch. 8). Classified information, for example, often hides not only what you have, but also what you know about the other. Similarly, credible signaling is not limited to conveying information about your capabilities and resolve, but also includes informing (or not) the other about what you know of him. Just as in the case of first-order intelligence, states have an incentive to misrepresent their beliefs about the other and to deceive them about what they know and what they do not know. Depending on their needs, they might exaggerate their perception of the other's capabilities, or on the contrary downplay them.

The explicit modeling of higher-order uncertainty leads to a more refined understanding of strategic interactions, with potentially important consequences in terms of the efficiency of bargaining. If states need to know what each other knows, then the conditions for peace are more stringent than the simple exchange of information about first-order attributes such as capabilities or resolve. More generally, explicitly modeling information structures may lead to a deeper understanding of states' strategies and more subtle implications of their communication. Future research will need to focus on the mechanisms by which states address the problems associated with higher-order uncertainty. International organizations, mediation, and monitoring, in particular, may all play a significant role in the creation of common knowledge.

References

Acharya, Avidit & Edoardo Grillo. 2015. "War with crazy types." *Political Science Research and Methods* 3(02):281–307.

Aumann, Robert J. 1976. "Agreeing to disagree." *The Annals of Statistics* 4(6):1236–1239.

Blainey, Geoffrey. 1988. *Causes of war.* Simon and Schuster.

Chadefaux, Thomas. 2011. "Bargaining over power: when do shifts in power lead to war?" *International Theory* 3(2):228–253.

Chadefaux, Thomas. 2012. "War as an Investment." Manuscript.

Crawford, Vincent P & Joel Sobel. 1982. "Strategic information transmission." *Econometrica* 50(6):1431–1451.

Farrell, Joseph & Robert Gibbons. 1989. "Cheap talk can matter in bargaining." *Journal of Economic Theory* 48(1):221–237.

Fearon, James D. 1994. "Domestic political audiences and the escalation of international disputes." *American Political Science Review* 88(3):577–592.

Fearon, James D. 1995. "Rationalist explanations for war." *International Organization* 49(3):379–414.

Fearon, James D. 1997. "Signaling foreign policy interests tying hands versus sinking costs." *Journal of Conflict Resolution* 41(1):68–90.

Ferris, John. 2007. *Intelligence and strategy: selected essays*. Routledge.

Fey, Mark & Kristopher W Ramsay. 2007. "Mutual optimism and war." *American Journal of Political Science* 51(4):738–754.

Filson, Darren & Suzanne Werner. 2002. "A bargaining model of war and peace: Anticipating the onset, duration, and outcome of war." *American Journal of Political Science* 46(4):819–837.

Fudenberg, Drew & Jean Tirole. 1991. *Game theory*. The MIT Press.

Garthoff, Raymond L. 2015. *Soviet leaders and intelligence: assessing the American adversary during the Cold War*. Georgetown University Press.

Gartzke, Erik. 1999. "War is in the error term." *International Organization* 53(3):567–587.

Harsanyi, John C. 2004. "Games with incomplete information played by Bayesian players, I–III: Part I. The Basic Model." *Management science* 50(12_supplement):1804–1817.

Holsti, Ole R. 1962. "The belief system and national images: A case study." *Journal of Conflict Resolution* 6(3):244–252.

Jervis, Robert. 1976. *Perception and misperception in international politics*. Princeton University Press.

Jervis, Robert. 2002. Signaling and perception: Drawing inferences and projecting images. In *Political psychology*, ed. Kristen Monroe. NJ: Lawrence Erlbaum: Hillsdale pp. 293–314.

Jervis, Robert. 2015. *Perception and misperception in international politics.* Princeton University Press.

Kahn, David. 2000. *Hitler's spies: German military intelligence in World War II.* Da Capo Press.

Kydd, Andrew H. 2005. *Trust and mistrust in international relations.* Princeton University Press.

May, Ernest R. 2014. *Knowing one's enemies.* Princeton University Press.

Mercer, Jonathan. 2010. "Emotional beliefs." *International Organization* 64(1):1–31.

Milbank, Dana. 2002. "Barriers to 9/11 inquiry decried. Congress may push commission." *Washington Post* p. A14.

Milgrom, Paul & Nancy Stokey. 1982. "Information, trade and common knowledge." *Journal of Economic Theory* 26(1):17–27.

Morris, Stephen. 1994. "Trade with heterogeneous prior beliefs and asymmetric information." *Econometrica* 62(6):1327–1347.

Morrow, James D. 1989. "Capabilities, uncertainty, and resolve: A limited information model of crisis bargaining." *American Journal of Political Science* pp. 941–972.

Morrow, James D. 2014. *Order within anarchy: The laws of war as an international institution*. Cambridge University Press.

Powell, Robert. 1996. "Bargaining in the shadow of power." *Games and Economic Behavior* 15(2):255–289.

Powell, Robert. 2004*a*. "Bargaining and learning while fighting." *American Journal of Political Science* 48(2):344–361.

Powell, Robert. 2004*b*. "The inefficient use of power: Costly conflict with complete information." *American Political Science Review* 98(2):231–241.

Prunckun, Hank. 2012. *Counterintelligence theory and practice*. Rowman & LIttlefield Publishers.

Reiter, Dan. 2003. "Exploring the bargaining model of war." *Perspective on Politics* 1(1):27–43.

Rubinstein, Ariel. 1989. "The electronic mail game: Strategic behavior under 'almost common knowledge." *The American Economic Review* 79(3):385–391.

Sartori, Anne E. 2002. "The might of the pen: A reputational theory of communication in international disputes." *International Organization* 56(1):121–149.

Schelling, Thomas C. 1980. *The strategy of conflict*. Harvard University Press.

Schultz, Kenneth A. 1998. "Domestic opposition and signaling in international crises." *American Political Science Review* 92(04):829–844.

Sebenius, James K & John Geanakoplos. 1983. "Don't bet on it: Contingent agreements with asymmetric information." *Journal of the American Statistical Association* 78(382):424–426.

Shore, Zachary. 2005. *What Hitler knew: the battle for information in Nazi foreign policy*. Oxford University Press.

Slantchev, Branislav L. 2003. "The principle of convergence in wartime negotiations." *American Political Science Review* 97(4):621–632.

Slantchev, Branislav L & Ahmer Tarar. 2011. "Mutual optimism as a rationalist explanation of war." *American Journal of Political Science* 55(1):135–148.

Smith, Alastair & Allan C Stam. 2004. "Bargaining and the nature of war." *Journal of Conflict Resolution* 48(6):783–813.

Snyder, Jack & Erica D Borghard. 2011. "The cost of empty threats: A penny, not a pound." *American Political Science Review* 105(3):437–456.

Toft, Monica Duffy. 2006. "Issue indivisibility and time horizons as rationalist explanations for war." *Security Studies* 15(1):34–69.

Trager, Robert F. 2010. "Diplomatic calculus in anarchy: How communication matters." *American Political Science Review* 104(2):347–368.

Van Evera, Stephen. 1999. *Causes of war: Power and the roots of conflict.* Cornell University Press.

Wagner, R. Harrison. 2000. "Bargaining and war." *American Journal of Political Science* 44(3):469–484.

Wittman, Donald. 1979. "How a war ends: A rational model approach." *Journal of Conflict Resolution* 23(4):743–763.

Yarhi-Milo, Keren. 2013. "In the eye of the beholder: How leaders and intelligence communities assess the intentions of adversaries." *International Security* 38(1):7–51.

Proofs

Proof of Lemma 1 (no separating equilibrium).

Consider first the possibility of a separating equilibrium in which A offers x^w upon observing (w_1, w_2) but x^s upon observing (s_1, s_2). B's posterior probabilities q_i of being at state s_i upon observing offer x^k and information set (θ_s) are then:

$$\begin{cases} q_1^B = q(w_1|x^w, (w_1)) = 1 \\ q_2^B = q(w_2|x^w, (w_2, s_1)) = 1 \\ q_3^B = q(s_1|x^s, (w_2, s_1)) = 1 \\ q_4^B = q(s_2|x^s, (s_2)) = 1 \end{cases}$$

First note that if there is to be a peaceful equilibrium, then it must be that

$$\begin{cases} p^w - c \leq x^w \leq p^w + c \text{ and} \\ p^s - c \leq x^s \leq p^s + c \end{cases}$$

Let $x^w = p^w - c + \alpha$ and $x^s = p^s - c + \beta$, where $0 \leq \alpha, \beta \leq 2c$. Clearly, given her beliefs, B always accepts A's offer. However, note that A^w's utility from playing (x^w, x^s) is $p^w - c + \alpha$. So A^w will deviate to an offer x^s instead of x^w

upon observing (w_1, w_2) if

$$q_1(p^w - c) + (1 - q_1)(p^s - c + \beta) > p^w - c + \alpha$$
$$\Rightarrow p^s - p^w > \frac{\alpha(p_1 + p_2)}{p_2} - \beta.$$

But we assumed that $p^s - p^w > \frac{2c(p_1+p_2)}{p_2}$, which is clearly greater or equal than $\frac{\alpha(p_1+p_2)}{p_2} - \beta$ for any $0 \leq \alpha, \beta \leq 2c$. This means that A will deviate (regardless of B's beliefs off the equilibrium), and hence there can be no separating equilibrium. □

Proof of Lemma 2 (pooling equilibrium).

Clearly there can be no pooling equilibrium in which A offers $x^* < p^s - c$, as a strong A would prefer war, and hence deviate. So consider a pooling equilibrium in which A offers $x^* = p^s - c$. In a pooling equilibrium, B's posteriors are dictated by Bayes' rule:

$$\begin{cases} q_1^B = q(w_1|x^*, (w_1)) = p_1 \\ q_2^B = q(w_2|x^*, (w_2, s_1)) = \frac{p_2}{p_2+p_3} \\ q_3^B = q(s_1|x^*, (w_2, s_1)) = \frac{p_3}{p_2+p_3} \\ q_4^B = q(s_2|x^*, (s_2)) = p_4 \end{cases}$$

B's beliefs off the equilibrium must be such that she assumes from any offer $x' \neq x^*$ that she is facing a weak A. If not, then at least A^w will want to deviate from offering x^*, and there can hence be no equilibrium in which both types of A offer x^*.

Consider now what B will do upon observing x^*. At (w_1), B will reject if $x^* = p^s - c > p^w + c$, which is assumed. At (s_2), B always accepts. Finally at (w_2, s_1), B accepts $x^* = p^s - c$ if the utility of accepting is greater than the utility of fighting weighted by the probability of the different types of A, i.e., if

$$\begin{aligned} 1 - x^* &\geq q_2^B(1 - p^w - c) + (1 - q_2^B)(1 - p^s - c) \\ \Rightarrow p^s - p^w &\leq \frac{2c(p_2 + p_3)}{p_2} \end{aligned}$$

which has been assumed.

Now does A have an incentive to deviate from this strategy? Clearly, a strong A will not deviate, given B's off the equilibrium path beliefs. But would a weak A deviate? A^w would deviate to offering $p^w + c$ (any higher offer will be rejected, and a lower offer is clearly not rational) if:

$$\begin{aligned} p^w + c &> q_1^A(p^w - c) + (1 - q_1^A)(p^s - c) \\ \Rightarrow \frac{2c(p_1 + p_2)}{p_2} &> p^s - p^w \end{aligned}$$

But $p^s - p^w > \frac{2c(p_1+p_2)}{p_2}$ was assumed, so A would not deviate and hence this constitutes an equilibrium. $\square$

Proof of Proposition 1.

The proof follows directly from lemma 3.1 and 3.2. To show that war occurs with positive probability in *every* PBE, simply replace $x^* = p^s - c$ in the proof of lemma 3.2 with $x^* = p^s - c + \gamma$, where $\gamma \in [0, 2c]$ is such that

$$\frac{2c - \gamma}{q_2^B} \geq p^s - p^c \geq \frac{2c}{q_2^A} - \gamma.$$