

Endogenous thresholds and assurance networks in collective action

Rationality and Society

2016, Vol. 28(2) 202–252

© The Author(s) 2016

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1043463115621523

rss.sagepub.com



Kris De Jaegher

Utrecht University School of Economics, The Netherlands

Abstract

This article treats a multi-player Stag Hunt where each player may have a different threshold (the number of other players that need to act along with the player for benefits of collective action to arise). Players are modeled as solving the strategic-uncertainty problem of whether or not to act, by assuring each other of their willingness to act. We show that in equilibrium there may, but need not, be homophily (players with the same thresholds seek assurance from each other) or a threshold-based social hierarchy (players with high thresholds, or “conservatives,” seek assurance from players with low thresholds, or “radicals,” but not vice versa). Put otherwise, a new strategic-uncertainty problem arises, namely, the problem of who should seek assurance from whom. We propose that players solve this problem by forming core-periphery assurance networks, with a number of players equal to the largest threshold in the core, and the remaining players in the periphery.

Keywords

Assurance, core-periphery assurance networks, endogenous thresholds, Stag Hunt

Introduction

Consider players who can benefit from collective action, such as voting for the same candidate, or participating in a riot (for other examples, see Granovetter, 1978). Each player prefers to act when all other players act,

Corresponding author:

Kris De Jaegher, Utrecht University School of Economics, Utrecht University,
Kriekenpitplein 21-22, 3584EC Utrecht, The Netherlands.

E-mail: k.dejaegher@uu.nl

and it is therefore possible for the players to achieve collective action. Yet, at the same time, if an insufficient number of other players act, the individual player does not want to act, and players may lock each other into inaction. While collective action is in all players' interests, acting is risky, because acting with too few players may come at a large cost. At the same time, not acting is safe, since it yields the status quo. There is a rationale both for players locking each other into collective action (as it is in everyone's interests) and for locking each other into inaction (as not acting is safe), and the players face *strategic uncertainty* (Brandenburger, 1996).

Players may counter strategic uncertainty by seeking assurance from each other that they will act (Kydd, 2000). Here, the fact that players are heterogeneous may play a role (Chwe, 2000). In particular, for "radical" players, the benefits of collective action may already arise when few other players act along with them, whereas for "conservative" players, these benefits only arise if many other players act, with "moderate" players somewhere in between. Each player may therefore have an individual *threshold* (Granovetter, 1978), that is, a number of other players that needs to act along with this player for benefits of collective action to arise to this player. The question arising then is who seeks assurance from whom, in what can be schematically represented as an *assurance network*. It then seems intuitive that radical players seek assurance only from each other, moderates seek assurance from each other and from radicals, and conservatives seek assurance from each other and from everyone else. Assurance networks would then be characterized by *homophily* (for an overview, see McPherson et al., 2001), where players with the same thresholds always seek assurance from each other, and by a *threshold-based social hierarchy*, where players with high thresholds seek assurance from players with lower thresholds, but not the other way around.¹

In this article, we develop a game-theoretic model, the *first* purpose of which is to correct these intuitions. We characterize the equilibria of our game, and show that the individual player need not preferentially seek assurance from other players with similar thresholds, thus contradicting homophily; moreover, it may be that low-threshold ("radical") players seek assurance from high-threshold ("conservative") players, but not vice versa, thus contradicting a threshold-based social hierarchy. The main insight here is to distinguish between a player's *exogenous threshold*, namely, how many players need to act for her to benefit from collective action, and her *endogenous threshold*, namely, how many players she requires assurance from in equilibrium. A player's endogenous threshold may exceed her exogenous threshold, and she may therefore behave more conservatively than would be expected from her exogenous threshold. It is not necessarily the

case that a player's exogenous threshold deterministically predicts how this player will behave, and where she will be positioned in any assurance network. On the contrary, a player's *embeddedness* (Granovetter, 1985) in a particular assurance network may determine how she behaves, and this behavior may be quite disconnected from her exogenous threshold.² In characterizing equilibrium assurance networks, we derive specific rules for the bounds within which a player's embeddedness in specific assurance networks may make her behave differently than would be expected by looking at her exogenous threshold.

The *second* purpose of this article is to show that the fact that the players seek assurance from each other, while resolving the strategic-uncertainty problem of whether or not to act, creates a new form of strategic uncertainty, namely, who should seek assurance from whom. We suggest a solution to this new strategic-uncertainty problem, by pointing out that a player who does not know the exogenous thresholds of the other players, or knows these thresholds but finds it cognitively challenging to adapt her strategy to the specific population she faces, can resort to an assurance architecture which works across a wide range of games, namely, the core-periphery architecture. In assurance networks with this architecture, a number of players equal to the maximum exogenous threshold in the population form the core, and all seek assurance from each other; the rest of the population (the periphery) seeks assurance from each player in the core, but not from each other.

Our prediction that players coordinate on core-periphery assurance networks is interesting in the light of the literature. Oliver and Marwell (1988) argue that a small core of radical players who act suffices as a critical mass (for an overview of critical mass theory, see Oliver and Marwell, 2001). Yet, Lohmann (1994) argues instead that a critical mass of acting players is only achieved if the core also includes moderate players, and illustrates this by means of the 1989 uprising in East Germany. Our model makes a similar prediction, but on different grounds. In Lohmann, moderates need to be included in the core because this provides convincing information about the desirability of an uprising to players not in the core. In our model, the many ways in which assurance networks can be formed for any given population, and the sensitivity of the set of equilibrium assurance networks to the population characteristics, creates strategic uncertainty on who should seek assurance from whom. As core-periphery architectures exist for a large class of games, our analysis suggests that players coordinate on these.

Before we explain how our article is structured, we point out that our entire analysis is rooted in non-cooperative game theory. Therefore, if we say that players achieve collective action, following non-cooperative game

theory, this does not refer to players having achieved this through some collective decision device but occurs because each player, given the behavior of the other players, found it in her individual interest to act. In the same manner, when we consider our players as forming a network, this is again in the manner of non-cooperative game theory (e.g. Bala and Goyal, 2000), where each player forms a link only when it is in her individual interest to do so. Also, while we consider coalitions of players to be able to achieve beneficial deviations from inefficient equilibria, following non-cooperative game theory, they only achieve this if each individual player's deviation is also a best response to the other players' deviations. Moreover, when we say that players coordinate, this refers to tacit coordination, entirely based on individual introspection and expectations. Finally, if we attribute to some players the epithet of leaders (and to other the epithet of followers), then the leaders are not to be interpreted as actively organizing collective action; simply, followers in our model only act conditional on leaders acting, but leaders may act even when followers do not act.

This article is structured as follows. Following a method of gradually increasing complexity, the section, "Preliminaries: From Stag Hunts to Assurance Games," starts with the standard two-player Stag Hunt, and gradually adds elements to this basic model which bring us closer and closer to our actual model, which is treated in the section, "The model: Heterogeneous Assurance Game," namely, the heterogeneous Assurance Game. The section, "Equilibria of heterogeneous Assurance Games," shows that the set of equilibria is large, even when allowing for equilibrium selection. The section, "Characteristics of core-periphery assurance architectures," points out some particular features of core-periphery assurance networks, namely, first that any Assurance Game may have a large number of equilibrium assurance networks with such an architecture, and second that many different Assurance Games have at least one assurance network with such an architecture. The section, "Core-periphery assurance architectures in asymmetric-information variants of the Assurance Game," shows that players who do not know others' exogenous thresholds, coordinate on core-periphery assurance networks. We end with a discussion in the section, "Discussion."

Preliminaries: From Stag Hunts to Assurance Games

We first introduce some simplified versions of our model, before we introduce the actual model in the section, "The model: Heterogeneous Assurance Game."

Table 1. Two-player Stag Hunt.

	Act	Don't act
Act	M, M	$0, -L$
Don't act	$-L, 0$	$0, 0$

Stag Hunt

The Stag Hunt is played by n players (for multi-player Stag Hunts, see (Carlsson and Van Damme, 1993a; Runge, 1984); the standard two-player Stag Hunt ($n=2$) is represented in Table 1. Each player decides whether or not to act without observing whether the other players do or do not act. A player who does not act always obtains the safe payoff 0, whatever the other players do. A player who acts obtains positive payoff M if all other players act as well, but obtains negative payoff $-L$ if at least one other player does not act. n is the player's *exogenous threshold*, namely, the number of players (including herself) who need to act before she obtains benefits from collective action. Depending on the size of n , we label players as radicals (small n), moderates (intermediate n), or conservatives (large n). All aspects of the game (available strategies, payoffs, rationality of the players) are common knowledge (each player knows all aspects of the game; each player knows that each player knows all aspects; etc.). A Nash equilibrium is a profile of strategies that are mutual best responses, that is, a situation in which no player wants to change what she does, given the behavior of the other players. The Stag Hunt has two pure-strategy Nash equilibria, namely, one where all players act and one where no player acts. The equilibrium where all players act is Pareto-efficient, meaning that if all players play this equilibrium, there is no alternative strategy profile where we can make one player better off without hurting the other players (for this reason, the equilibrium where neither player acts is Pareto-inefficient).³

The fact that there are multiple Nash equilibria creates *strategic uncertainty* (Brandenburger, 1996) to the players. They may consider that the collective action equilibrium is Pareto-dominant (i.e. is the unique Pareto-efficient equilibrium), which could be argued to create mutual expectations that this equilibrium will be played (Harsanyi and Selten, 1988). Yet, the individual player may at the same time consider that acting is a risky strategy. If the cost of acting alone is large, even limited doubt that other players do not act may lead the individual player not to act. Thus, mutual expectations that the joint inaction equilibrium will be played cannot be excluded (i.e. the joint inaction equilibrium is risk dominant;⁴ Harsanyi and Selten, 1988). In order to achieve the collective action equilibrium, on top of all

aspects of the game being common knowledge, it must also be common knowledge among the players that the collective action equilibrium will be played (Aumann and Brandenburger, 1995).

Two modeling assumptions of the Stag Hunt deserve attention. *First*, contrary to the Prisoner's Dilemma (cf. Olson, 1965; Tucker, 1950), the Stag Hunt has a Nash equilibrium where collective action is achieved. For this reason, among others, Runge (1984), Skyrms (2004), and Centola (2013) have argued for the use of Stag Hunt games rather than Prisoner's Dilemmas to model collective action. *Second*, no player gets a benefit when acting by herself, meaning that all players have an exogenous threshold n . For $n=2$, suppose instead that one player i would have an exogenous threshold of 1. Then it would be optimal for this player to act, whatever the other player j does. Given this fact, it would be a best response for player j , who continues to have exogenous threshold 2, to act as well; there would now be a unique Nash equilibrium where both players act. Achieving collective action is a problem in the Stag Hunt because a *critical mass* of acting players needs to be achieved for any benefits of collective action to arise.

Heterogeneous Stag Hunt

We now allow for heterogeneous exogenous thresholds, so that radical, moderate, and conservative players may all interact in the same game. Player i now has exogenous threshold t_i , meaning that she obtains payoff M when acting only if at least $(t_i - 1)$ other players act as well. We assume that for each player, $t_i \geq 2$, so that each player considers collective action (i.e. action with more than one player) as necessary for benefits of acting to arise. We further assume that $t_i \leq n$, ensuring that a Nash equilibrium exists where each player acts.⁵

In any heterogeneous Stag Hunt, pure-strategy Nash equilibria where all players act and where no player acts continue to exist. On top of this, depending on the distribution of exogenous thresholds across players, many other Nash equilibria may exist where only a strict subset of the players acts. Specifically, it is easy to see that for any heterogeneous Stag Hunt where a subset of t players has exogenous threshold t or lower, but where no player has exogenous threshold exactly equal to $(t + 1)$, a Nash equilibrium exists where only t players act. Put otherwise, the more gaps there are in the exogenous threshold distribution, the more Nash equilibria there are. Each time, such a gap presents a hurdle to the players on the higher side of the gap, as these players need to jointly act to obtain benefits from collective action. In the absence of a gap in the exogenous thresholds, a subset of t players that acts in equilibrium creates a *bandwagon effect* of extra players who find it a best response to act. If t players

act in equilibrium, then any player with exogenous threshold $(t+1)$ will find it a best response to act. Given this fact, any player with threshold $(t+2)$ will also act, and so on.⁶

As is clearly seen, in any Nash equilibrium where moderates act, radicals must act as well; and in any Nash equilibrium where conservatives act, moderates and radicals must be acting as well. Yet, radicals may act without moderates acting, and moderates may act without conservatives acting. Technically, if player i with exogenous threshold t_i finds it a best response to act in equilibrium, then the same must be true for each player with exogenous threshold t_i or smaller. While such a result is intuitive, it begs the question of how players achieve collective action in the first place. Whatever their exogenous thresholds, the problem of strategic uncertainty continues to exist for all players: given that even radicals do not benefit from acting by themselves, and find it costly to act alone, they may equally doubt whether a sufficient number of people will act. In the next section, "Assurance Game," we introduce a model of how players may reassure each other of their willingness to act.

Assurance Game

Let us now revisit the Stag Hunt, and turn it into an Assurance Game.⁷ Looking more closely at the source of strategic uncertainty in the Stag Hunt, the problem of the individual player i who considers acting is that, even if i thinks that the probability is high that j holds beliefs that collective action can be achieved and that j acts based on these beliefs, i may still attach positive probability to j instead thinking that collective action is unachievable, and to j not acting. As long as the cost L of acting with too few players is sufficiently large, player i now still decides not to act. The players therefore may only achieve collective action if they are somehow able to reassure each other that they will act.

We introduce such assurance into the homogeneous Stag Hunt, and turn it into an Assurance Game in the following way. We assume that a majority of players is willing to act, in being trustful that collective action can be achieved. Trustful players are thereby assumed to find the Pareto dominance of the collective action equilibrium salient, and if all players would be trustful in this way, they would always achieve collective action. The problem is that a minority of players is not willing to act. These players may be considered as focusing on the fact that the joint inaction equilibrium is risk dominant. Without further information about the type of players she is facing, even a trustful player will not act. Yet, we assume that at a cost, trustful players are able to identify each other, and in this way may achieve assurance that they can safely act.⁸

In particular, *first*, we formalize the typical player i 's consideration that any other player j may either believe or not believe that collective action is achievable, by assuming that there are two types of players, namely, willing players and unwilling players. Specifically, we assume that with (small) probability ε , the individual player is in state u (unwilling), and never acts. To make such behavior of unwilling players consistent, we assume that they obtain payoff $-L$ whenever they act, and payoff 0 whenever they do not. With the complementary probability $(1 - \varepsilon)$, the player is in state w (willing), and has the same payoffs as in the homogeneous Stag Hunt. We assume that a willing player will always act when common knowledge is achieved with the other players that they are all in state w . Yet, we also assume that, given that $\varepsilon > 0$, L is large enough for a willing player not to act as long as not knowing that every other player is willing as well, that is, $(1 - \varepsilon)M - \varepsilon L < 0$. *Second*, we model each player as being able, at a cost c , to find out the state of an individual other player. Intuitively, the individual player may put in effort to check for cues that other players are willing to act. When player i checks the state of player j , using a network metaphor, we represent this graphically as a directed link $i \rightarrow j$ from a node with label i to a node with label j .⁹ We further assume that $(1 - \varepsilon)^{n-1}M - (n-1)c > 0$, so that the individual willing player who expects that each other willing player checks the state of each other player and acts if each other player is willing prefers to check all other players' states and act if they are all willing, to not checking any states of other players and not acting. Given our assumption, $(1 - \varepsilon)M - \varepsilon L < 0$, collective action only takes place if each willing player indeed checks the states of all other players.

Formally, the Assurance Game proceeds as follows. At stage 1, Nature independently for each player decides with probability $(1 - \varepsilon)$ that she is willing and with probability ε that she is unwilling. Each player observes her own state, but not the state of the other players. Also, each player decides for each other player whether or not to check this player's state. We assume that the individual player is not able to observe whether individual other players are checking her state, or the states of other players. Also, when player i checks player j 's state, she finds out player j 's state, but not the states of the players which player j may herself be checking (i.e. information is local). At stage 2, players simultaneously decide whether or not to act. Given that the Assurance Game is a game of incomplete information, we need a variant of the Nash equilibrium concept, namely, the perfect Bayesian equilibrium (henceforth PBE; Fudenberg and Tirole, 1991). A PBE is similar to the Nash equilibrium, but with the addition that players form beliefs in accordance with Bayesian updating, and that these beliefs are confirmed in equilibrium.¹⁰

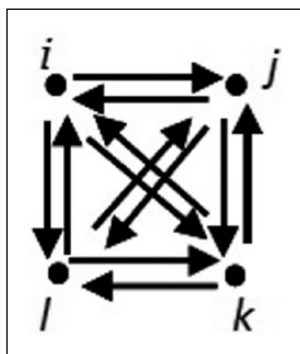


Figure 1. Complete assurance network for four-player game.

Given that $(1-\varepsilon)M - \varepsilon L < 0$ and $(1-\varepsilon)^{n-1}M - (n-1)c > 0$, the Assurance Game has two PBEs, namely, one where each willing player checks the state of each other player and acts when each other player is also willing. Additionally, the game has a PBE where no player checks the state of any other player and where no player acts. In network terms, any PBE can be represented as an *assurance network*. In such an assurance network, an arrow from player i to player j denotes that player i , when in state w , checks the state of player j ; if player i forms links to a set of other players, then this means that player i in state w only acts if all players to which she forms a link are also willing. The Assurance Game therefore has two PBE assurance networks, namely, the complete network (where each player has a directed link to each other player) and the empty network. The complete assurance network for the example of an Assurance Game with four players is depicted in Figure 1. This complete network nicely represents that players only act when achieving common knowledge that they are all in state w : each directed path in the PBE assurance network represents one of the knowledge statements that need to be achieved in order to achieve common knowledge. Note that this common knowledge is achieved not because the individual player directly observes each willing player observing that each other player is willing. As information is local, the individual player only directly observes the states of all the players of which she checks the states. Yet the individual willing player in equilibrium believes that all other willing players have also checked all other players' states.

Multiple equilibria thus continue to exist in the Assurance Game, and we move from a problem on whether or not to act in the Stag Hunt, to a problem on whether or not to check other's states in the Assurance Game. A player in

state w who checks the state of another player, and finds her to be in state w as well, may still doubt whether the other player also checked the states of other players. Yet, we assume that players in state w can resolve this problem (which justifies calling them willing players), by introducing the equilibrium selection concept of *information⁺-proofness*. A formal definition of this concept will be given below in Definition 1, when we treat heterogeneous Assurance Games. The reasoning is that the empty PBE assurance network will not be played, because each willing player realizes that she can become better off by checking the state of the other players, and by acting after having achieved common knowledge that each player is willing; moreover, she expects other players to reason in the same way when in state w . Contrary to what is the case in the PBE, an equilibrium is information⁺-proof if it is not only the case that each individual player does not become better off by unilaterally deviating, but it is also the case that sets of players do not become better off by jointly deviating (where deviations are mutually best responses).¹¹

We draw attention to several aspects of the Assurance Game. *First*, what is strategic uncertainty (=uncertainty about what other players will do) in the Stag Hunt is formalized as structural uncertainty (=uncertainty about the payoffs of other players) in the Assurance Game (see Brandenburger, 1996, for this distinction). While strategic uncertainty is intuitively present in the original Stag Hunt, its source is at the same time not clear when the other player can only be of a single type. Even though we add an element to the Stag Hunt that was originally not there, namely, a subset of players who never act, the advantage is that we are able to formalize the source of the individual player's uncertainty.¹² *Second*, we do not model assurance as signaling one's own willingness to others, but as paying attention to the cues of other's willingness. Even if players would send signals about their willingness (cf. Kim and Sobel, 1996), paying attention to such signals also comes at a cost, and is a strategic decision, without which signaling cannot be effective (see Binmore and Samuelson, 2001 for this argument). Given its importance, we focus purely on the decision of whether or not to pay attention. We also assume that players are not able to observe from each other whether they are paying attention—otherwise the mere fact of incurring the cost of paying attention may become a signal of one's willingness to act (cf. Spence, 1973). *Third*, the manner in which players in the Assurance Game are able to achieve collective action bears some resemblance to the secret-handshake argument in evolutionary models (Robson, 1990): players can safely act because they are able to recognize some trait in each other (for a similar argument, see Güth and Kliemt, 1994, 1998). The main difference between our model and this literature is that in the latter players can

change their traits (so, unwilling players can become willing), whereas in our model, we consider the traits as given. *Fourth*, the concept of information⁺-proofness assumes that willing players realize that it is in their joint interest to deviate from an equilibrium with joint inaction. Why not simply apply this concept to all players in the Stag Hunt? In this manner, the collective-action equilibrium is immediately predicted, without any need to turn the Stag Hunt into an Assurance Game. Yet, the point here is that we would then again not be able to reflect the intuition that players of the Stag Hunt face strategic uncertainty, as this strategic uncertainty would be resolved without any further action taken by the players.

So far, the contribution of turning Stag Hunts into an Assurance Game may seem limited. We are simply arguing that most players should trust that collective action can be achieved, and that if these players are able to recognize each other, they will coordinate on collective action. This leads to a single predicted equilibrium where each willing player checks whether or not all other players are willing as well, and acts when all other players are willing. As we will now show, once we consider heterogeneous Assurance Games, it is no longer clear who should seek assurance from whom.

The model: Heterogeneous Assurance Game

We finally come to the class of games of interest, namely, the class of the heterogeneous Assurance Games. Closest in the literature is Chwe's (2000) threshold game, with the difference that in our model, networks are not given exogenously but are formed strategically, and that our players have an exogenous threshold as well as a state. Contrary to the homogeneous Assurance Game previously treated (all players have exogenous threshold n), each player i may now have *any* exogenous threshold t_i such that $2 \leq t_i \leq n$. At stage 1, independently for each player and independently from her exogenous threshold, Nature decides with probability $(1 - \varepsilon)$ that she is in state w , and with probability ε that she is in state u . We initially assume that players are able to observe each other's exogenous thresholds (an assumption that we will modify in the section, "Core-periphery assurance architectures in asymmetric-information variants of the Assurance Game"). Yet, we continue to assume that they do not observe each other's states, and that each player can at a cost c check the state of any other player (where again players do not observe whether or not other players are checking them or others, and do not observe the states of players checked by other players). At stage 2, the players simultaneously decide whether or not to act.

We now justify in more detail the modeling assumptions. *First*, why assume that players can be in states w or u , separately from their exogenous

thresholds? Do the exogenous thresholds not already reflect the extent to which they are willing? The relevance of the states is that, whatever their exogenous thresholds, players face strategic uncertainty: just as the radical players in the two-player homogeneous Assurance Game, and the more conservative players in a multi-player homogeneous Assurance Game, players will continue to face strategic uncertainty when radicals and conservatives all interact in the same game. Yet, is it not plausible that players with a lower exogenous threshold (i.e. more radical players) are more likely to be in state w ? The point here is that as long as there is positive probability that a radical is unwilling, and as long as the cost of acting with too few players is large, attaching a separate probability ε_i of being in state u to a player with exogenous threshold t_i , where ε_i is smaller the smaller t_i , does not make any difference for our results. It continues to be the case that a player in state w with exogenous threshold t_i checks the states of at least $(t_i - 1)$ other players.

Second, another plausible way in which players with different exogenous thresholds may differ from each other lies in their costs of paying attention to other players' states. Specifically, it is plausible that players find it cheaper to check the states of players who have the same exogenous threshold, simply because such players may be socially closer. This would seem to naturally lead to homophily, where players with the same exogenous thresholds seek assurance from each other. Yet, as long as it is not the case that it is prohibitively costly for the individual player to check the state of a player with an exogenous threshold that is very different, who checks whom is a matter of coordination, and players in equilibrium need not check the states of the players for which checking is the cheapest.

Third, in the homogeneous Assurance Games treated above, as all players have the same exogenous threshold, it makes sense to consider the players as simultaneously deciding whether or not to act. By assumption, each player in state w then obtains benefits from acting only when $(n-1)$ other players currently act as well, or have previously acted; because of this, no subset of players wants to act first. Yet, in a heterogeneous Assurance Game, it makes sense for more conservative players to act before other players do. Indeed, sociologists have criticized game theorists for focusing too much on games where players decide whether or not to act simultaneously, rather than sequentially (Granovetter, 1978: 1434; Macy, 1991: 730–731). Nevertheless, if there are gaps in the exogenous threshold distribution, assurance continues to be necessary. It is straightforward but cumbersome to develop a variant of the game where players may either at a cost check the states of other players, or may alternatively wait and see whether other players have previously acted. As the results for such a variant of the game are very similar, we focus on games where players always decide simultaneously whether or not to act.¹³

We continue to assume that players in state u obtain payoff $-L$ whenever they act, and obtain payoff 0 whenever they do not act, where we now assume that this is the case whatever their exogenous thresholds. A player in state w with exogenous threshold t_i who acts obtains payoff M if at least $(t_i - 1)$ other players act, and obtains payoff $-L$ otherwise; she obtains payoff 0 whenever she does not act. We assume that

$$(1 - \varepsilon^{n-1})M - \varepsilon^{n-1}L < 0 \quad (1)$$

so that a willing player with exogenous threshold 2 who has not checked any of the other players' states prefers not to act even if all other players are expected to act as soon as they are in state w . This ensures that a necessary condition for a willing player with *any* exogenous threshold t_i to act is that she checks the states of at least $(t_i - 1)$ other players, and finds at least $(t_i - 1)$ of these players to be willing. We further assume that

$$(1 - \varepsilon)^{n-1}M - (n - 1)c > 0 \quad (2)$$

so that a willing player with exogenous threshold n who expects that all willing players will act prefers to check the states of all other players, to not checking any states and not acting. Given this condition, players with lower exogenous thresholds also prefer checking to not checking. We finally impose a condition so that a player always checks the minimal number of other players necessary to be able to achieve certainty that the other player will act. In particular, suppose that checking the states of $\tau_i - 1$ players is the minimal number of other players a player i in state w needs to check in order to achieve certainty that a sufficient number of other players will act, then condition (3) imposes that player i will not check the states of more other players than necessary. The right-hand side of condition (3) puts an upper limit on the payoff that can be achieved by checking the states of x additional players, by assuming that it suffices that any subset of $(\tau_i - 1)$ players in the set of $(\tau_i - 1 + x)$ players who are all in state w suffices to achieve collective action with certainty

$$(1 - \varepsilon)^{\tau_i - 1}M - (\tau_i - 1)c > \sum_{y=(\tau_i - 1)}^{\tau_i - 1 + x} \binom{\tau_i - 1 + x}{y} (1 - \varepsilon)^y \varepsilon^{\tau_i - 1 + x - y} M - (\tau_i - 1 + x)c \quad (3)$$

To see that conditions (1)–(3) are compatible, consider first ε approaching zero. Then condition (2) becomes $M - (n - 1)c > 0$, and condition (3) is always valid. It follows that if $M - (n - 1)c > 0$, for sufficiently small (but

positive) ε , conditions (2) and (3) are valid. Taking now such a small but positive ε , we see that condition (1) is valid as long as L is sufficiently large.

Equilibria of heterogeneous Assurance Games

Structure of PBE assurance architectures

Given assumptions (1)–(3), Lemma 1 specifies that in any PBE of a heterogeneous Assurance Game where a (weak or strict) subset of players act, each willing player with exogenous threshold t_i who acts checks the states of a number $(\tau_i - 1)$ of other players equal to at least $(t_i - 1)$, and acts only if *each* of these players are in state w . Therefore, the loss of acting with too few players is so high that our typical player does not want to gamble by checking a lower number of other players than is necessary to convince herself that her exogenous threshold will be achieved. Moreover, given the strategies of the other players, each willing player in equilibrium checks the minimum number of players necessary to establish that payoff M will be obtained with probability 1. Thus, at the same time, the probability of any given player being in state u is sufficiently low, for the individual player never has to check a higher number of other players than is necessary to convince herself that her exogenous threshold will be achieved.

Lemma 1. Given assumptions (1)–(3), in any PBE of a heterogeneous Assurance Game:

- (i) players in state u do not check the states of other players, and do not act;
- (ii) any player i in state w with exogenous threshold t_i
 - a. checks the states of a number $(\tau_i - 1)$ (with $\tau_i \geq t_i$) of other players, and only acts when each of these other players is also in state w ;
 - b. chooses other players to check such that, given the strategies of the other players, $(\tau_i - 1)$ is the minimum number of players of which she needs to check the states, in order to achieve a belief of 1 that at least $(t_i - 1)$ other players will act.

The proof of Lemma 1, and of all lemmata and propositions in the article, is found in Appendix 1. Given Lemma 1, we can again concisely represent any PBE of a heterogeneous Assurance Game as a PBE assurance network. Again, if a player i has links to a set of other players in a PBE assurance network, this means that in the corresponding PBE, player i when in state w only acts if each of the players in the set is willing as well; if player i does not have any links in a PBE assurance network, it means

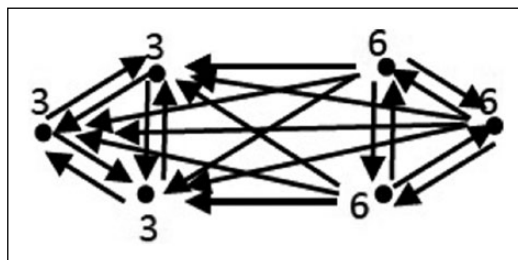


Figure 2. PBE assurance network for game with exogenous thresholds (3, 3, 3, 6, 6, 6).

that she does not act. In a heterogeneous Assurance Game, a PBE assurance network can typically have many more architectures than just the complete network or the empty network. For example, consider a simple game with six players, where each time exactly three players have exogenous thresholds 3 and 6; put otherwise, we have three radicals, and three conservatives. In the PBE assurance network in Figure 2, the radicals now only need assurance from each other, and not from the conservatives.

As a first step toward characterizing the PBE assurance networks of heterogeneous Assurance Games, we say that two PBE assurance networks (of the same game, or of two different games) have the same *assurance architecture*, if they are identical when purely seen as a set of nodes with links between them, without players assigned to the nodes. When there exists at least one heterogeneous Assurance Game that has a PBE assurance network with a specific given assurance architecture, then we call this assurance architecture a *PBE assurance architecture*.

In Proposition 1, we now list the characteristics of the set of all PBE assurance architectures, aggregated across all heterogeneous Assurance Games we consider (namely, those where each player's exogenous threshold ranges from two to at most n). The formulation of Proposition 1 requires the introduction of some graph-theoretical concepts. A *clique* is a maximal subset of nodes in a graph which all form directed links to each other (note that one-player cliques are also possible).¹⁴ Graphically, we represent a clique by a circle, and connect any pair of nodes i and j inside the circle by a single line, as short-hand notation for two directed links in both directions between i and j . For instance, in the assurance architecture in Figure 3, there are two cliques. A *directed clique link* $x \rightarrow y$ from a clique x to a clique y represents in short-hand a set of directed links from *every* node in clique x to *every* node in clique y . For example, in Figure 3, the arrow from the right clique to the left clique represents that every node in the right clique has a directed link to every node in the left clique. A *directed clique graph* is a partition of a set of

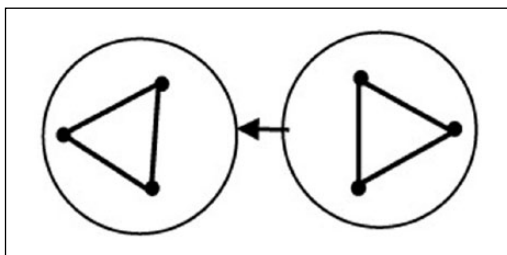


Figure 3. Assurance architecture with leading clique and follower clique.

nodes into cliques, where cliques may or may not be connected by directed clique links. An *isolated clique* is a clique without any directed clique links to or from other cliques; a *non-isolated clique* has at least one directed clique link to or from another clique. A *directed clique path* is a series of cliques and clique links such that $v \rightarrow x \rightarrow y \rightarrow \dots \rightarrow z$, a *directed clique cycle* is a directed clique path such that $v \rightarrow x \rightarrow y \rightarrow \dots \rightarrow z \rightarrow v$.

The *length of a directed clique path* is the number of directed clique links on this clique path. The *longest directed clique path* between a clique x and a clique y is the directed clique path with the largest length between these two cliques. When the longest directed clique path from clique x to clique y has length 1, we say that clique x is a *direct follower* of clique y . An *acyclic directed clique graph* is a directed clique graph not containing any directed clique cycles. By a basic graph-theoretic result, every acyclic directed clique graph contains at least one *leading clique* (=a clique from which no directed clique link departs), and at least one *end clique* (=a clique at which no directed clique link arrives).¹⁵

Proposition 1 shows that necessary and sufficient conditions for a graph to be a PBE assurance architecture are that (i) it is an acyclic directed clique graph,¹⁶ in which (ii) if a clique x is a direct follower of a clique y , who is in turn a direct follower of a clique z , ... and so on, then clique x also has directed clique links to cliques y, z, \dots ; (iii) every leading clique contains at least two nodes; and (iv) the number of isolated one-player cliques is either zero, or is at least two.

To understand the results in Proposition 1, consider first the intuition for any PBE assurance network to take on the architecture of an acyclic directed clique graph. Players in a leading clique z all form links to each other, because for players to act it is not only important that they know that a number of other players are in state w , but also that they know what these other players know. Therefore, the players in a leading clique achieve (local) common knowledge that they are all in state w . For additional players outside of the leading clique to also act, given that there may be gaps

in the exogenous threshold distribution, intuitively these players will also only act after having formed a clique of their own. These players will then not only form links to each other, but also to each player in the leading clique, and thus form a direct follower clique y . Because the players in the leading clique z only act if they achieve common knowledge that they are all in state w , every player in the follower clique y forms a link to every player in the leading clique z . By the same reasoning, the direct follower clique y may itself have a direct follower clique x . Given that the players in clique y only act if all players in clique z are in state w , the players in clique x will not only form links to every player in clique y , but also to every player in clique z and so on (Proposition 1(ii)). A PBE assurance network cannot contain any directed clique cycle, because then by the same reasoning, each player contained in the directed clique cycle would only act if all players contained in the clique cycle are in state w , and so all players contained in the directed clique cycle would have to be in one and the same clique.

Furthermore, given that each player's exogenous threshold is assumed to equal at least 2, each leading clique z in a PBE assurance network should include at least two players (Proposition 1(iii)). Finally, it cannot be that there is only a single player who does not form any links (Proposition 1(iv)). To see why, if $(n-1)$ players form links, leaving one isolated one-player clique, then these must each have exogenous threshold $(n-1)$ or smaller. Given that we have assumed that no player's exogenous threshold can exceed n , the remaining player must have exogenous threshold n or lower. But then, this player when in state w benefits individually from connecting to a weak subset of the rest of the players, and the isolated clique cannot be maintained.

To see that the conditions in Proposition 1 are also sufficient, note that for any acyclic directed clique graph with the properties listed in Proposition 1, one can easily find a heterogeneous Assurance Game with an exogenous threshold distribution such that this graph is a PBE assurance network for the given game. Simply, attach to each node in the graph with at least one link a player with an exogenous threshold equal to the number of links formed, plus one; furthermore, attach to each node without links a player with exogenous threshold equal to the number of nodes with at least one link, plus two. In this manner, one is able to enumerate all PBE assurance networks where cliques are characterized by homophily and by a threshold-based social hierarchy. For example, for the assurance architecture in Figure 3, one finds a heterogeneous Assurance Game for which a PBE assurance network has this architecture, by assigning players with exogenous thresholds 3 to each node in the leading clique, and players with exogenous thresholds 6 to each node in the follower clique, as depicted in Figure 4.¹⁷

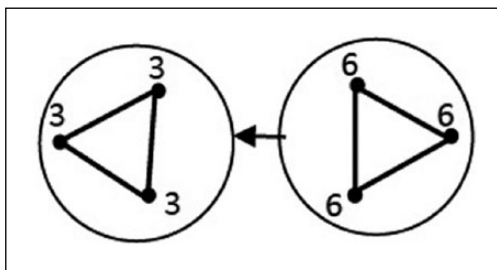


Figure 4. PBE assurance network for game with exogenous thresholds (3, 3, 3, 6, 6, 6).

Proposition 1. Necessary and sufficient conditions for a graph with n nodes to be a PBE assurance architecture are that:

- (i) the graph is an acyclic directed clique graph;
- (ii) for any clique v on a directed clique path $v \rightarrow x \rightarrow y \rightarrow \dots \rightarrow z$ of the graph, clique v forms a directed clique link to all cliques x, y, \dots, z ;
- (iii) each leading clique in the graph has cardinality of at least two; and,
- (iv) the graph contains either no, or at least two isolated one-player cliques.

Following Chwe (2000), any PBE assurance architecture can literally be interpreted as a social hierarchy in the following sense. In any PBE assurance architecture, along a directed clique path ending in a leading clique z , consider the direct follower clique y of the leading clique z , the direct follower clique x of the direct follower clique y , and so on. Then the players in the leading clique may be considered as “leaders,” in the sense that no other player along the mentioned directed clique path acts, if they do not act. In the hierarchy, the direct follower clique y of the leading clique z is the next in rank, in that players in this clique only act if all leaders act, but in that all other players along the mentioned directed clique path only act if the players in clique y all act and so on. It should be stressed that reference to social roles such as “leaders,” “followers,” and “followers of followers” does not refer to these players acting sequentially, but rather to the fact that all players along the directed clique path check the states of the leaders, and all non-direct followers of the leaders check the states of the leaders and of the direct followers of the leaders, and so on. Players continue to act simultaneously.

In general, along an acyclic directed clique path, the longer the longest directed clique path from a clique x to a clique y , the lower the rank of clique y compared to clique x , meaning that in the corresponding PBE

assurance network, the players in clique y are less likely to act compared to those in clique x . In this sense, an acyclic directed clique graph can be interpreted as a hierarchy of cliques.

Homophily and a threshold-based social hierarchy

The proof of Proposition 1 shows that to any assurance architecture with the properties listed, corresponds a PBE assurance network of a specific Assurance Game. This is done by constructing networks characterized by *homophily* and by a *threshold-based social hierarchy*.¹⁸ We now show that these properties are not generally valid for PBE assurance networks.

A *first* reason for this is that there may simply not be enough players with identical exogenous thresholds to guarantee homophily and a threshold-based social hierarchy. For example, consider the Assurance Game with exogenous thresholds (2, 3, 3, 5, 5, 5). This game has a PBE assurance network with the three lowest threshold players assigned to the leading clique, two of the threshold-5 players assigned to a direct follower clique of the leading clique, and the remaining threshold-5 player assigned to a one-player direct follower clique of the two-player follower clique, as depicted in Figure 5. While the player with exogenous threshold 2 could achieve collective action when mutually checking her state with another player with exogenous threshold 2, only players with higher exogenous thresholds are available. Let the threshold-2 player consider only checking the state of a threshold-3 player. The problem now is that, at best, this player will not only want to check the state of the threshold-2 player, but also of the other threshold-3 player. It follows that the threshold-3 player, when in state w , will only act if both the threshold-2 player and the other threshold-3 player are also in state w . Given this fact, the threshold-2 player is forced to check the states of both threshold-3 players. The behavior of the threshold-2 player becomes indistinguishable from a player with exogenous threshold 3. This is why we say that the threshold-2 player forms an *endogenous threshold* equal to 3.

By the same reasoning, the threshold-5 player in the one-player follower clique would be able to do with fewer links if there were a fourth threshold-5 player; together with this player, she could then form a second two-player clique that is a direct follower of the leading clique. Yet, such a player is not available, and she cannot assure herself that at least a number of players equal to her exogenous threshold act, by connecting only to the leading clique. Connecting to one player in the two-player direct follower clique of the leading clique is not sufficient, because she can only ascertain that this player acts when the other player in the clique also acts. She is therefore

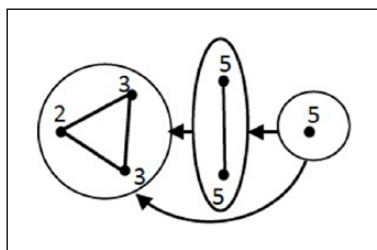


Figure 5. PBE assurance network for game with exogenous thresholds (2, 3, 3, 5, 5, 5).

forced to connect to all other players and to form an endogenous threshold equal to 6. While she has the same exogenous threshold as the other threshold-5 players, she has a lower rank in the social hierarchy, and is less likely to act than them, so that there is in this instance no systematic threshold-based social hierarchy.

Second, even when a PBE exists where there is homophily and a threshold-based social hierarchy, players may play a Pareto-inefficient PBE where these properties are violated. For instance, for the case of six players, consider an Assurance Game with exogenous thresholds (2, 2, 3, 4, 5, 5). This game has the PBE assurance network depicted in Figure 6(a). This is in line with Oliver and Marwell (1988) who in the context of a sequential model argue that as soon as a small core of radical players act, in a bandwagon effect, less radical players will join in. Yet, the specified Assurance Game also has the PBE assurance network depicted in Figure 6(b), which violates both homophily and a threshold-based social hierarchy.¹⁹ For each individual player in the leading clique, if all others expect that any collective action can only take place if all players in the leading clique are in state w and all check each other's states, the best response of the individual player in the leading clique is to follow the equilibrium strategy. Note that the players with exogenous thresholds lower than 5 are forced to set an endogenous threshold of 5, because of the presence of threshold-5 players in the leading clique. The threshold-2 player in the follower clique would as such like to check the state of a single player in the leading clique. But, as such, a player only acts when finding out that all other players in the leading clique are in state w , the threshold-2 player in the follower clique is forced to check their states as well, and set an endogenous threshold of 6. Note that the threshold-2 player in the follower clique has a lower rank than the threshold-5 players in the leading clique and is less likely to act than them. The example in Figure 6(b) may be seen as reflecting *embeddedness* (Granovetter, 1985): it is not only the case that a player's inherent characteristics (namely, her exogenous threshold) determine her position in the assurance network; it may

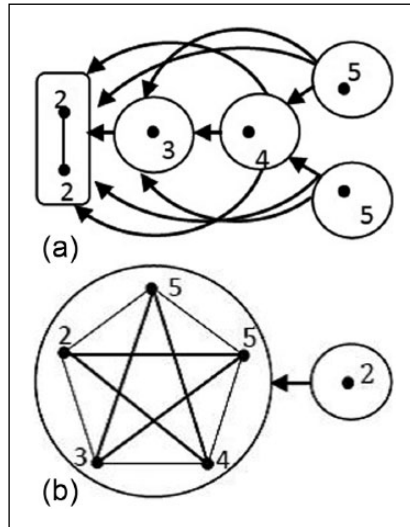


Figure 6. Two PBE assurance networks for game with exogenous thresholds (2, 2, 3, 4, 5, 5).

oppositely be the case that her embeddedness in a specific assurance network determines her apparent characteristics (in the form of her endogenous threshold). From an empirical point of view, this suggests that it is hazardous to adopt a revealed preference approach, and infer an agent's degree of conservativeness from her behavior in collective action problems.

Third, players' mutual expectations may lock them into not acting, which may be true for all, or for a subset of players. For instance, in the case with exogenous threshold distribution (2, 2, 4, 4, 4, 4), a PBE exists where each willing threshold-2 player forms a link to the other threshold-2 player, and acts when the other player is also willing, but where the threshold-4 players do not form links and never act (Figure 7(a)). Moreover, a PBE exists where none of the players form links and none of them act (Figure 7(b)).

Fourth, players' mutual expectations may lock them into forming an excess of links. Again for the same exogenous threshold distribution, an example is found in Figure 8(a). The clique link from the follower clique to the leading clique is redundant. Yet, if all threshold-4 players expect from each other that they will each only act when checking whether both threshold-2 players are in state w , then it is a best response for the individual threshold-4 player to check this. The most extreme case of such a mechanism is a PBE where each player checks the state of each other player, where such a PBE exists for any of the exogenous threshold distributions we consider. Simply, if each player expects each other player to check the states of all

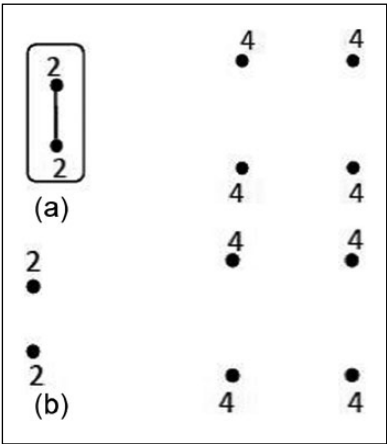


Figure 7. Two PBE assurance networks including non-acting players for game with exogenous thresholds (2, 2, 4, 4, 4, 4).

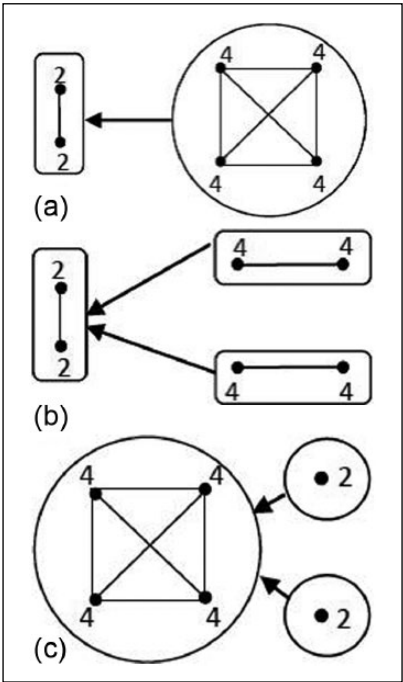


Figure 8. Three PBE assurance networks for game with exogenous thresholds (2, 2, 4, 4, 4, 4); (a) is not information^{+/-} proof, and not Pareto-efficient; (b) is information^{+/-} proof, and Pareto-efficient; (c) is information^{+/-} proof, but not Pareto-efficient.

other players, and to only act when all other players are in state w , it is a best response for the individual player to follow the same strategy.

Structure of information^{+/-}-proof PBE assurance architectures

These arguments showing that PBE assurance networks need not be characterized by homophily or a threshold-based social hierarchy, at the same time, illustrate that any given Assurance Game may have many PBEs. In order to select among the multiple PBEs, we first apply the concept of information⁺-proofness (Definition 1), already introduced for the homogeneous Assurance Game. The argument here is that players cannot get stuck into a situation where only a strict subset of the players act. Intuitively, non-acting players when in state w should then be able to get information on the states of other players and of each other, and still achieve collective action.

Definition 1. For an Assurance Game, consider any PBE assurance network g . Then this PBE assurance network g is *information⁺-proof*, if it is *not* possible for any weak subset of players N_x to jointly add links to g such that

- (i) adding these links strictly increases the expected payoff of *each* player in N_x ;
- (ii) the links that each individual player in N_x adds are part of a best response for this player, given the links that all the other players in N_x add.

By Lemma 1, a player or group of players who in a PBE assurance network forms links, cannot become better off by forming further links, as this would only make it less likely for them to act, and thereby decreases their expected payoffs. Thus, the joint deviations considered in Definition 1 are only relevant for players not currently forming any links. Moreover, given the restriction that the joint deviation must consist of mutual best responses, by Proposition 1, the joint deviation should involve the players in the set N_x partitioning themselves and forming cliques, and possibly connecting these cliques by means of clique links to cliques in g and/or by means of clique links within N_x .

Second, we add the assumption that if a subset of players can do better by collectively forming fewer links than they currently do, they will do this. Formally, this leads to the definition of the concept of an information⁻-proof PBE in Definition 2. Information⁻-proofness excludes PBE assurance networks where players lock each other into checking an excess of states, such

as the extreme case where each player checks the state of each other player, and only acts when each other player is willing.

Definition 2. For an Assurance Game, consider any PBE assurance network g . Then this PBE assurance network g is *information⁻-proof*, if it is *not* possible for any weak subset of players N_x to jointly remove from g links departing from players in the set N_x , such that

- (i) removing these links weakly increases the expected payoff of *each* player in N_x ;
- (ii) the remaining links that each individual player in N_x maintains continue to be part of a best response for this player, given the links that all the other players in N_x maintain.

Given the restriction that the remaining links of players jointly deviating by removing links must be mutual best responses, by Lemma 1 and Proposition 1, the effect of information⁻-proofness is that, starting from a given PBE assurance network, if a subset of players can get better off by splitting up an existing clique into subcliques, and/or by deleting some of their current clique links (resulting in a new PBE assurance network), they will do so.

We finally define the concept of an information^{+/-}-proof PBE, which is simply a PBE that is both information⁻-proof and information⁺-proof. Note that this means that a PBE assurance network is information^{+/-}-proof if a subset of players cannot get better off by either jointly adding links, or by jointly removing links; not considered are joint deviations where a subset of players both jointly adds links, and jointly removes links (we justify this below).

Definition 3. A PBE assurance network of the heterogeneous Assurance Game is *information^{+/-}-proof* if it is both information⁺-proof (Definition 1) and information⁻-proof (Definition 2).

Having already characterized the set of all PBE assurance architectures in Proposition 1, we now characterize the set of all *information^{+/-}-proof PBE assurance architectures*. An information^{+/-}-proof PBE assurance architecture is any assurance architecture for which at least one Assurance Game exists that has an information^{+/-}-proof PBE assurance network with this architecture. As shown in Proposition 2, any information^{+/-}-proof PBE assurance architecture has the same characteristics as listed in Proposition 1, with the additional characteristic that such an architecture does not contain isolated one-player cliques.²⁰

Proposition 2. Necessary and sufficient conditions for a graph with n nodes to be an information^{+/-}-proof PBE assurance architecture are conditions

(i)–(iv) in Proposition 1, with the additional condition that the graph does not contain isolated one-player cliques (condition (v)).

For any given Assurance Game, how does the set of information^{+/−}-proof PBE assurance networks compare to the set of Pareto-efficient PBE assurance networks? To provide an answer, we start by looking at some characteristics of the latter set. First, given assumption (2), it is clear that in any Pareto-efficient PBE, all players act. By assumption (2), even if the individual player in state w can only assure herself that any other player acts when finding out that each other player is in state w , this still leaves the player better off than not checking the states of any other player, and not acting. In the set of PBEs where all players act, one PBE is then Pareto-superior to another PBE because the Pareto-superior PBE saves at least one player linking costs and increases her probability of achieving collective action, whereas the situation of the remaining players stays the same. Second, for any Assurance Game with an exogenous threshold distribution such that at least one PBE assurance network exists with the property that all players' endogenous thresholds equal their exogenous thresholds, *all* Pareto-efficient PBE assurance networks have this property, and are all Pareto-equivalent. Note that this property implies that the Pareto-efficient PBE assurance networks are then characterized by both homophily and a threshold-based social hierarchy. Yet, it is not the case for all Assurance Games that the set of Pareto-efficient PBE assurance networks has this property. For instance, consider a four-player game with exogenous threshold distribution (2, 3, 3, 3). Then in any Pareto-efficient PBE assurance network, we have a leading clique of three players, and a follower clique of one player. Each individual player is better off being in the leading clique, and players prefer different Pareto-efficient PBE assurance networks, where three of these networks violate both homophily and a threshold-based social hierarchy.

The set of information^{+/−}-proof PBE assurance networks of a given Assurance Game is typically strictly larger than the set of its Pareto-efficient PBE assurance networks. For instance, the assurance network in Figure 8(c) is information^{+/−}-proof, because starting from this network, a subset of players either adding links, or a subset of players deleting links, does not make any such subset of players better off. Yet, the network is not Pareto-efficient, because in the assurance network in Figure 8(b), the threshold-4 players are equally well off (as they achieve collective action when all in state w , and incurring the same linking costs), whereas the threshold-2 players are better off (as they incur lower linking costs and are more likely to achieve collective action).

If the concept of information^{+/−}-proofness does not lead us to eliminate all Pareto-inefficient PBEs, then why not simply adopt an equilibrium selection criterion whereby if a deviation to a Pareto-superior PBE

assurance network requires a subset of players to both add and remove links, the players are able to achieve this? We argue that such switches may be too complex to perform for players. For instance, in the example of Figure 8, a switch from the Pareto-inefficient PBE assurance network in Figure 8(c) to the Pareto-efficient one in Figure 8(b) requires a complete reversal of social roles, where the leaders become followers, and the followers become leaders. With information^{+/−}-proofness, subsets of players can add an additional social role to an established hierarchy if this is in their mutual interests, or in an existing hierarchy, a single existing social role can be split up in several subroles if this is in the interest of these players. Yet, more complex changes in the hierarchy are assumed not to be feasible to the players.²¹

Our concept of information^{+/−}-proofness bears resemblance to equilibrium concepts employed in the game-theoretic literature modeling the formation of social and economic networks, which allows players to jointly deviate, but puts some limits on the joint deviations that are allowed (for an overview, see Bloch and Jackson, 2006). For example, the static concept of pair-wise stability (Jackson and Wolinsky, 1996) assumes that a pair of players will mutually form a link when it is in both their interests, while a link can be unilaterally deleted if this is in one player's individual interest.²² The limit put here on the sort of joint deviations that players can make is that only two players can jointly deviate. Players constrained to such joint deviations may be seen as myopic, as it may take several rounds of jointly adding links and unilaterally deleting links to come to a Pareto-superior equilibrium (Jackson and Watts, 2002). In our model, concentrating on joint deviations by pairs of players leads to limited results, since players may have exogenous thresholds in excess of 2. Our Pareto-improving joint deviations may involve any number of players, but we assume the players to be constrained in that they may only either jointly add links or jointly delete links, but not both.²³

Closest related to information^{+/−}-proofness is Chwe's (2000) concept of minimal sufficient networks. These are assurance networks that are sufficient to ensure that all players act, but do so in such a way that this result cannot also be obtained by deleting links from the assurance network. An essential difference is that in our model, links are formed strategically.

Assignment of exogenous thresholds to (information^{+/−}-proof) PBE assurance architectures

We now characterize the set of all (information^{+/−}-proof) PBE assurance networks. As a first step, we show that any Assurance Game we consider has at least one (information^{+/−}-proof) PBE assurance network. Indeed, while in

Propositions 1 and 2 we derived necessary and sufficient conditions on the structure of an (information^{+/-}-proof) PBE assurance architecture, this does not show that each Assurance Game has at least one such assurance network. This is shown in Proposition 3.

Proposition 3. Every heterogeneous Assurance Game has at least one (information^{+/-}-proof) PBE assurance network.

We next show that for any given heterogeneous Assurance Game, a plethora of PBEs typically exists, even if we limit the analysis to information^{+/-}-proof PBEs. We show this not by enumerating all the (information^{+/-}-proof) PBEs for each Assurance Game, but rather, we “reverse engineer” by taking any (information^{+/-}-proof) PBE assurance architecture as characterized in Propositions 1 and 2, and by deriving three rules that allow one to enumerate all heterogeneous Assurance Games that have a (information^{+/-}-proof) PBE assurance network with the given architecture.²⁴ The fact that this is typically a large number of games shows that oppositely the typical heterogeneous Assurance Game has many (information^{+/-}-proof) assurance networks. The rules are derived in three technical propositions in Appendix 2. By means of examples, we provide the intuitions for these rules, which separately look at isolated one-player cliques, non-isolated one-player cliques, and multi-player cliques.

The *first* rule (Proposition A1) concerns the assignment of players with specific exogenous thresholds to isolated one-player cliques in PBE assurance architectures. An example is found in Figure 9. In PBE assurance networks, the exogenous thresholds r , s , and t in the one-player cliques may each independently be either 5 or 6. Intuitively, in order for players in one-player cliques not to want to connect individually, there must be a gap between the endogenous thresholds in the connected cliques, and exogenous thresholds of the players in the isolated one-player cliques, witnessed here by the fact that the players in the one-player cliques may not have exogenous threshold 4 or lower. Note that Figure 9 can never be information⁺-proof, as not every player forms links. Therefore, the first rule only applies to non-information⁺-proof PBE assurance networks.

The *second* rule (Proposition A2) concerns the assignment of players with specific exogenous thresholds to non-isolated one-player cliques in PBE assurance architectures. As the decisions in one-player cliques depend on individual behavior, the second rule applies equally whether or not we allow for non-information^{+/-}-proof PBE assurance architectures. The rule is simple. Consider in any given PBE assurance architecture the number of links l that depart from a node i in a one-node clique x , and consider the largest number of links l' strictly smaller than l that may alternatively be

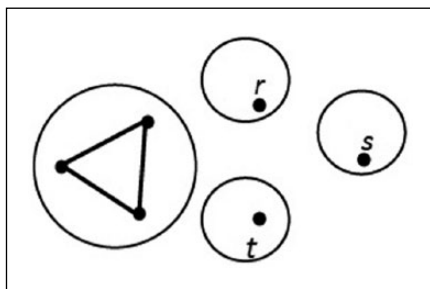


Figure 9. PBE architecture with one three-player clique and three one-player cliques; assignment of exogenous thresholds to isolated one-player cliques.

formed from i to another set of cliques (and in accordance with the condition in Proposition 1(ii)). Then the exogenous threshold t_i for a player assigned to node i can have any value such that $(l' + 2) \leq t_i \leq (l + 1)$.

As a first example, consider the player with exogenous threshold r in Figure 10(a), where this player forms $l = 5$ links. The alternative is to form only $l' = 2$ links to the leading clique. This means that $4 \leq r \leq 6$. Indeed, when player i has exogenous threshold 3 or lower, she will only connect to the leading clique. However, when player i has exogenous threshold 4, she is forced to connect to at least one player in the three-player clique. But as she can only be sure that one of these players acts if they all act, she connects to all of them.

As a second example, consider the player with exogenous threshold s in Figure 10(b), where this node has $l = 4$ links. The alternative is to form only $l' = 2$ links to the leading clique. This means that $4 \leq s \leq 5$. As a third example, consider the player with exogenous threshold t in Figure 10(b), where this node has $l = 5$ links. Then one can see that $l' = 4$, which applying the rule means that $6 \leq t \leq 6$.

These examples show that there is more flexibility to assign exogenous thresholds to players in non-isolated one-player cliques, the larger the cliques of which such cliques are direct followers. Of interest is then to see to what extent the exogenous threshold t_i attached to the one-player clique x can be lower than the endogenous thresholds (and therefore also lower than the exogenous thresholds) in the clique y of which it is a direct follower—thereby violating a threshold-based social hierarchy. Suppose that l' is simply equal to the number of links that players in clique y form to other players outside of the clique, and that clique y has n_y players. Then the endogenous threshold of each player in clique y equals $(l' + n_y)$. Given that it must be that $(l' + 2) \leq t_i \leq (l + 1)$, it follows that the exogenous threshold of the player in the one-player clique can at most be equal to the endogenous thresholds

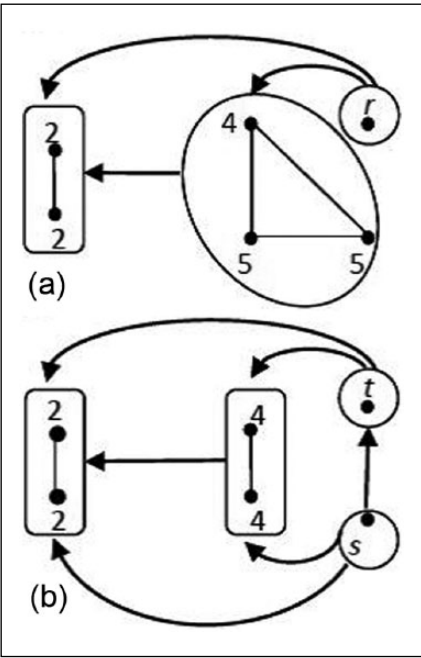


Figure 10. PBE assurance architectures for two six-player games; assignment of exogenous thresholds to non-isolated one-player cliques.

in the clique y of which it is a direct follower, if clique y contains two players, and can be strictly lower than the endogenous thresholds in the clique y only if clique y contains three or more players. For example, in Figure 10(a), the endogenous thresholds in the three-player clique are each time equal to 5, but the exogenous threshold in the direct follower clique may be lower at 4. In Figure 10(b), the endogenous thresholds in the two-player follower clique are each time 4, and the exogenous threshold in the direct follower clique of this two-player clique is at least 4.

The *third* rule (Proposition A3) concerns the exogenous thresholds of players in multi-player cliques of PBE assurance architectures. As an example, in the PBE assurance architecture in Figure 11, we look at which exogenous thresholds players in the four-player clique can have. Note first that no player can have exogenous threshold 3, as otherwise she would individually prefer to connect only to the players in the leading clique. However, in a PBE assurance network, the exogenous threshold of any player in the four-player clique may range from four to six. In an information^{+/−}-proof PBE assurance network, however, there may not be more than one threshold-4 player, and no more than two threshold-5 players—but otherwise,

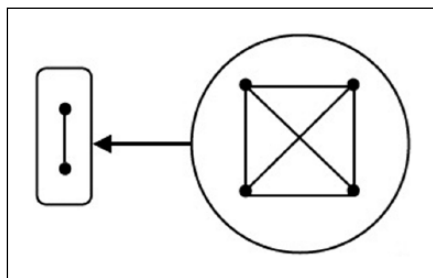


Figure 11. PBE assurance architecture for six-player game; assignment of exogenous thresholds to multi-player cliques.

anything goes. It is clear then that there is most scope for assigning exogenous thresholds to nodes in a multi-player clique, if it is a leading clique (as the exogenous thresholds in the multi-player clique are then not constrained by the links they already form outside of the clique), and if the clique is large (simply because the exogenous threshold can then take on larger values). In particular, in a leading clique, there may not be more than one threshold-2 player, not more than two threshold-3 players, and so on. In other words, if the leading clique has s players, there must be at least two threshold- s players, at least three players with exogenous threshold $(s-1)$ or higher, at least four players with exogenous threshold $(s-2)$ or higher, and so on.

The second and third rules allow us to look more generally at the scope for violations of homophily and of a threshold-based social hierarchy. By the second rule, it is only possible that a player in clique y has a weakly lower exogenous threshold than the endogenous threshold in clique z of which it is a direct follower, if clique z is a multi-player clique, and clique y is a one-player clique. By the same rule, if clique z is a multi-player clique, and clique y is a one-player clique, in a direct-follower clique x of clique y , the exogenous thresholds must be strictly higher than the exogenous threshold in clique y . Moreover, the exogenous thresholds of players in clique x must also be strictly higher than the endogenous threshold in clique z , as players in clique x would otherwise not individually want to link to players in clique y . It follows that there can at best be local violation of a threshold-based social hierarchy. Furthermore, consider the general rule that a player with exogenous threshold 2 can at most be at distance one from a leading clique, and that a player with exogenous threshold t_i ($t_i \geq 3$) can at most be at longest directed distance $(t_i - 2)$ from a leading clique (thus, a threshold-3 player can be at most at distance 1 from a leading clique, a threshold-4 player at most at distance 2, etc.). This shows that the lower the players' exogenous thresholds, the closer they need to be to the leading clique. Yet,

at the same time, any player can be positioned in a leading clique, as long as the leading clique contains sufficiently many players. By the third rule, because of her embeddedness in a large clique, a player with a low exogenous threshold (i.e. an inherently radical player) may form a high endogenous threshold (i.e. may act conservatively). Yet, note that oppositely, a player with a high exogenous threshold can never be made to form a low endogenous threshold (i.e. an inherently conservative player can never behave radically).

In conclusion, as to the typical PBE assurance architecture correspond multiple Assurance Games that have an information^{+/-}-proof PBE assurance network with this architecture, the typical Assurance Game also has multiple information^{+/-}-proof PBEs. This shows that, while assurance solves the strategic-uncertainty problem of whether or not to act, it creates a new strategic-uncertainty problem, namely, who should seek assurance from whom. In the next two sections, “Characteristics of core-periphery assurance architectures” and “Core-periphery assurance architectures in asymmetric-information variants of the Assurance Game,” we argue that players can solve this problem by coordinating on an assurance network with a core-periphery architecture.

Characteristics of core-periphery assurance architectures

We continue our analysis by treating two characteristics of *core-periphery* PBE assurance networks, namely, networks with a single leading clique containing exactly t_{\max} players, and any other remaining players in one-player cliques that are direct followers of the leading clique. At an intuitive level, these characteristics suggest that players may coordinate on such assurance networks. The section, “Core-periphery assurance architectures in asymmetric-information variants of the Assurance Game,” develops extensions of our model where this is indeed the case.

Typically Assurance Games have many PBE assurance networks with a core-periphery architecture

A majority of the PBE assurance networks of any given Assurance Game may have the core-periphery architecture. As an example, consider the Assurance Game with exogenous threshold distribution (2, 2, 3, 4, 4, 4). This game has 11 information^{+/-}-proof PBE assurance networks, of which seven have the core-periphery architecture. To see this, note first that the network has one bandwagon information^{+/-}-proof PBE assurance network,

where the two threshold-2 players are in a leading clique, and all other players are in one-player cliques. Second, it has three information^{+/-}-proof PBE assurance networks with the two threshold-2 players again in a leading clique, but with a clique of two threshold-4 players as a direct follower of this leading clique; three such networks are obtained because there are three ways to put the three threshold-4 players in such a follower clique. Third, it has seven information^{+/-}-proof PBE assurance networks with a leading clique of four players, and all other players in one-player direct follower cliques. To see this, note that there are six ways to construct a leading clique containing exogenous thresholds (2, 3, 4, 4) and one way to construct a leading clique containing exogenous thresholds (3, 4, 4, 4).

In general, as the analysis in the previous section, “Equilibria of heterogeneous Assurance Games,” shows, among all PBE assurance architectures that may exist for a given Assurance Game, the core-periphery architecture is the most flexible for the assigning of players to social roles. This is because, as follows from the previous section, players’ exogenous thresholds are most flexible, first, in large leading cliques, and second, in one-player cliques that are direct followers of a large clique.

PBE assurance networks with a core-periphery architecture exist across a large set of Assurance Games

A large class of Assurance Games with different exogenous threshold distributions may have (information^{+/-}-proof) PBE assurance networks with a core-periphery architecture—even though otherwise the set of (information^{+/-}-proof) PBE assurance networks compared across these Assurance Games may look very different. Formally, Proposition 4 shows that the class of Assurance Games with identical number of players and identical maximal exogenous threshold all share PBE assurance networks with the same core-periphery architecture. A subclass of this class of Assurance Games, where additionally the exogenous threshold distribution is sufficiently biased toward higher exogenous thresholds, shares information^{+/-}-proof PBE assurance networks with the same core-periphery architecture. This result follows directly by applying the third rule treated in the section, “Equilibria of heterogeneous Assurance Games,” to derive the exogenous thresholds that can be assigned to the nodes in a leading clique with t_{\max} players, and the second rule to the exogenous thresholds that can be assigned to one-player follower cliques. Because, as previously noted, the assignment of exogenous thresholds is most flexible in large leading cliques, and in one-player follower cliques, it is intuitive that PBE assurance networks with the core-periphery architecture exist across a wide class of Assurance Games.

Proposition 4. Consider a core-periphery assurance architecture consisting of a leading clique containing t_{\max} players, and $(t_{\max} - n)$ one-player direct-follower cliques of the leading clique. Then, at least one assurance network with this architecture

- (i) is a PBE assurance network for *any* Assurance Game;
- (ii) is an information^{+/-}-proof PBE assurance network for any Assurance Game with at least two players with exogenous threshold t_{\max} , at least three players with exogenous threshold $(t_{\max} - 1)$ or higher, at least four players with exogenous threshold $(t_{\max} - 2)$ or higher, and so on.

Core-periphery assurance architectures in asymmetric-information variants of the Assurance Game

The section, “Characteristics of core-periphery assurance architectures,” showed that the typical Assurance Game may have many (information^{+/-}-proof) PBE assurance networks with the core-periphery architecture, and that moreover a large class of Assurance Games with different exogenous threshold distributions shares (information^{+/-}-proof) PBE assurance networks with the same core-periphery architecture. Intuitively then, players who face strategic uncertainty about whom to seek assurance from, may put their bets on an assurance network with a core-periphery architecture, because most PBEs in any given Assurance Game may take this form, and because this architecture works across a large set of Assurance Games. This section formalizes this intuition by considering two variants of the Assurance Game with asymmetric information about the exogenous thresholds. In both variants, players coordinate on a core-periphery assurance network, where the core contains a number of players equal to the largest exogenous threshold. Intuitively, while following Oliver and Marwell (1988), a small core of radical players who reassure each other may suffice to create a bandwagon effect, the individual player may doubt about the endogenous thresholds that other players may form (formalized as asymmetric information in this section). A safe strategy is to assume that each player behaves as conservatively as the most conservative player in the population. Therefore, our model explains collective conservatism, as predicted by Kuran (1988).

It should be noted that in both extensions below, strategic uncertainty continues to exist. While only core-periphery assurance networks survive, each strict subset of t_{\max} players may be allocated to the core, so that many equilibria continue to exist (especially since there may be many core-periphery PBE assurance networks). Yet, all remaining PBE assurance

networks have exactly the same architecture, and players may further coordinate using a randomization device to assign a random subset of t_{\max} players to the core, and the rest of the players to the periphery. In this way, each of the core-periphery PBE assurance networks is played with equal probability, and each player obtains the same ex ante expected payoff.²⁴

Asymmetric information: Players do not know other players' exogenous thresholds

We first treat a variant of the heterogeneous Assurance Game where players do not know each other's exogenous thresholds. Stages 1 and 2 proceed as in the section "The model: Heterogeneous Assurance Game," but are now preceded by an additional stage 0. At stage 0, Nature chooses an exogenous threshold t_i for each player i according to the density function $f(t_i)$, where $\sum_{t_i=2}^{t_{\max}} f(t_i) = 1$. Each player observes her own exogenous threshold, but not those of the other players. n , t_{\max} , and the density function are all common knowledge.

We maintain the assumptions in equations (2) and (3). Proposition 5 shows that under a modified version of equation (1), implying a large cost L of acting with too few players, every information^{+/−}-proof PBE assurance network has the same architecture, consisting of a leading clique containing exactly t_{\max} players, with the other players are connected to it in one-player cliques. Intuitively, as long as L is sufficiently large, as the probability that all other players have exogenous threshold t_{\max} is positive, players only want to act if collective action is ensured even if all players have the maximal exogenous threshold.

Proposition 5. Under assumptions (1)–(2), a sufficiently large L can be found such that every information^{+/−}-proof PBE assurance network of the Assurance Game with asymmetric information about the exogenous thresholds consists of a leading clique containing exactly t_{\max} players, with any remaining players in one-player direct follower cliques of the leading clique.

Players can check the exogenous threshold distribution at a cost

In the spirit of Mengel (2012) and Weesie et al. (2009), each individual may be involved in many Assurance Games, each time involving other players with a possibly different exogenous threshold distribution.²⁶ Even if they observe the endogenous threshold distribution, cognitively constrained players may find it costly to re-adjust their strategies to each individual Assurance Game, and would therefore benefit from adapting one and the same strategy across multiple games.

We model this as follows. We maintain the stage 0 added in the previous section, “Asymmetric information: Players do not know other players’ exogenous thresholds,” but now assume that at a cost k , each player at stage 0 can additionally decide to observe all the exogenous thresholds of the other players. Thus, each player faces the binary decision whether to observe at a cost the exogenous threshold of each other player, or whether not to observe others’ exogenous thresholds. All other modeling assumptions are maintained.

Clearly, an equilibrium exists where no player checks the exogenous thresholds of others. Simply, if no other player checks the exogenous thresholds, it is not in the interest of the individual player to do so, as other players do not act upon their exogenous thresholds anyway. The information^{+/–}-proof PBEs described in Proposition 4 continue to exist, with the added feature that the description of each strategy includes each player’s decision not to check any other player’s state. As long as the cost of checking the exogenous threshold distribution is sufficiently large, these are also the Pareto-efficient equilibria.

Even if equilibria where no player checks the exogenous threshold distribution are Pareto-efficient, equilibria may also exist where players lock each other in each checking the threshold distribution. In such equilibria, players are then assumed to coordinate on a different PBE assurance network for each exogenous threshold distribution that may occur (where each such PBE assurance network is in line with the analysis in the section, “Core-periphery assurance architectures in asymmetric-information variants of the Assurance Game”). While it is technically straightforward to describe such equilibria, an implicit assumption is then that players are able to coordinate on a specific PBE assurance network for each separate exogenous threshold distribution. Yet, at stage 0, before deciding on whether or not to check the exogenous threshold distribution, the individual player may consider it as uncertain which PBE assurance network players will coordinate on for each separate exogenous threshold distribution. The description of equilibria where players check the exogenous threshold distribution therefore does not take into account that strategic uncertainty may make it less attractive to check the exogenous threshold distribution. The fact that there may be strategic uncertainty for each exogenous threshold distribution that the players may observe, gives further credence to equilibria where players do not check the exogenous threshold distribution.²⁷

Discussion

Players attempting to achieve collective action face strategic uncertainty about whether or not other players will act. We construct a model where

players solve this problem by assuring each other about their willingness to act. The *first* message of our article is that such assurance seeking creates a new form of strategic uncertainty, namely, who should seek assurance from whom. Though players may have different exogenous thresholds determining how many other players should act before it is in their individual interest to act, the thresholds that players form endogenously (i.e. the number of other players they need assurance from before they decide to act) may exceed their exogenous thresholds. While players' exogenous thresholds in part determine their position in any equilibrium assurance network, and therefore also determine their behavior, oppositely the way in which players happen to be positioned in assurance networks determines their behavior, apart from their exogenous thresholds. For example, when a radical player happens to seek assurance from conservative players, the radical player's behavior may become indistinguishable from that of a conservative player. It is because of the many ways in which a player can be embedded into an assurance network, each time possibly making the player behave in a different way, that the mentioned new problem of strategic uncertainty (who should seek assurance from whom) arises.

The *second* message of our article is that players may solve this new strategic-uncertainty problem by forming core-periphery assurance networks, where a core of players equal to the largest exogenous threshold seeks assurance from each other, and where any remaining players seek assurance only from all players in the core. Such players only achieve collective action when there is consensus about the desirability of collective action, as each player only acts if the most conservative player acts. Intuitively, if players do not know each other's exogenous thresholds, the strategic-uncertainty problem is largely resolved, as the safe strategy to each player is to seek assurance as if all other players are conservatives. Since not knowing other players' exogenous thresholds solves the strategic-uncertainty problem, it makes sense that players' ignore each other's exogenous thresholds, and act as if each player is a conservative as the most conservative player in the population.

Acknowledgements

I would like to thank Stephanie Rosenkranz, Bastian Westbrock, and two anonymous referees for helpful comments.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Such a threshold-based social hierarchy is found in so-called bandwagons (Granovetter, 1978; Yin, 1998). In a simple three-player setting, a radical acts by herself. Observing this, the moderate acts as well. Finally, observing that both the radical and the moderate acted, the conservative acts as well. Note that in such a bandwagon, players physically observe other players having acted before, whereas in our model, they seek assurance from each other that they will act in the near future.
2. The point that individuals are shaped by the networks in which they are situated, is made more generally by Lazer (2001).
3. Additionally, the Stag Hunt has a mixed equilibrium where players randomize between acting and not acting in such a way as to keep each other indifferent between acting and not acting, and therefore willing to randomize in the specified way. Just like the join inaction equilibrium, the mixed equilibrium is Pareto-inefficient, and for simplicity, we limit the analysis to pure-strategy equilibria.
4. Suppose that player i expects player j to act with probability p , and not to act with probability $(1-p)$, then player i finds acting better than not acting iff $pM - (1-p)L > 0$ iff $p(M+L) > L$ iff $p > L/(M+L)$. Note that $L/(M+L) > 1/2$, given that $L > M$.
5. Agents with exogenous thresholds higher than n may be considered as not playing the game.
6. We stress that this interpretation does not refer to players deciding to act, having physically observed that other players are already acting (as in Granovetter, 1978), but is instead fully based on introspection of the players: our players continue to make their decisions simultaneously.
7. The earliest reference on the Stag Hunt (Sen, 1967) uses the term Assurance Game as synonymous to the Stag Hunt. In this article, we instead use the term Assurance Game to refer to a Stag Hunt augmented with active assurance seeking by the players. For an alternative model of assurance, where players assure each other by cooperative gestures, see Kydd's (2000) Reassurance Game.
8. From the perspective of the literature on risk dominance, willing players are identical to the players in the Stag Hunt, and therefore would equally well consider the joint inaction equilibrium as risk dominant. No player would therefore check the state of another player. Moreover, given that we add to the model unwilling players, players would be even less inclined to act. This is the approach in the literature on global games (Carlsson and Van Damme, 1993b), where strategic uncertainty is resolved by adding a degree of structural uncertainty to the model, in favor of risk dominant equilibria.
9. This contrasts with Chwe (1999, 2000), where links take the form of player i communicating her threshold to player j . Our links thus run in the opposite direction of those in Chwe, and refer to player i checking player j 's state.
10. In the context of the game-theoretic literature on social and economic network literature (for an overview, see Jackson, 2010), separate literatures have looked at strategic network formation, where no game is played on the network (e.g.

Bala and Goyal, 2000; Jackson and Wolinsky, 1996), and at games played on exogenously given networks (e.g. Bramoullé and Kranton, 2007; Chwe, 2000). This article is one of a smaller number of articles that looks simultaneously at network formation and at the game played on this network (see e.g. Goyal and Vega-Redondo, 2005). Specific about our model is that the game played on the network is a global interaction game, rather than the local interaction game that is usually treated in the literature (see e.g. Corbae and Duffy, 2008 for local Stag Hunt games), and that the links of the network do not determine who interacts with whom but determine who gets information from whom.

11. The notion of information⁺-proofness as described here fits the concept of a strong Nash equilibrium (Aumann, 1959), where no individual player finds it beneficial to deviate unilaterally, and no group of players finds it beneficial to deviate jointly. Yet, note that joint inaction equilibrium in the Prisoner's Dilemma is then not a strong Nash equilibrium, as each player benefits from a joint deviation toward collective action—and there does not appear to be a specific reason then to consider Stag Hunts rather than Prisoner's Dilemmas. However, collective action is not enforceable, since the individual player would again want to deviate. The concept of a coalition-proof Nash equilibrium (Bernheim et al., 1987) limits joint deviations to enforceable deviations (no individual or group of players wants to deviate from the joint deviation). The joint deviations in non-information⁺-proof PBEs are enforceable, and thus the concept of information⁺-proofness lies closest to the concept of coalition-proofness.
12. The approach taken here resembles literature that investigates the effect of players with small probability being of a behavioral type, which always behaves in the same way, into games, as initiated by Kreps and Wilson (1982).
13. See also Footnote 17. More remote are models by Bikhchandani et al. (1992), Lohmann (1994), and Sunstein (2009), where players who see that other players are already acting, also find out information on their *own* benefits and costs of acting. In these models, it is possible that players are influenced to behave more radically. In our model, on the contrary, players can only be influenced to behave more conservatively.
14. In graph-theoretic terms, what we refer to as a clique is a *maximal* clique. Unless stated otherwise, we use the term clique to denote a maximal clique for shortness.
15. See Bang-Jensen and Gutin (2009: 32), Proposition 2.1.1. In graph theory, an end node is a node that receives a link but from which no node departs. In what we call an end clique, no directed clique link arrives but at least one directed clique link departs.
16. Chwe (2000) studies assurance networks in a non-strategic way. An assurance network is sufficient if it allows for collective action of all players, and is minimal sufficient if it does not contain a subgraph that is also sufficient. Minimal sufficient networks may have overlapping cliques. To see why, consider an Assurance Game with four players labeled 1 to 4 who all have threshold 3. Consider a network with bilateral links between 1 and 2, 1 and 3, 2 and 3, 2 and

- 4, and 3 and 4. This network, which consists of overlapping cliques, is minimal sufficient, because it does not contain a subgraph that also allows all players to act. In our analysis, however, this is not a PBE assurance network, because the presence of the cycle $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$ means that the four players should be in the same clique.
17. Figure 4, and the examples that follow in the rest of the article, also represent PBEs of a variant of the Assurance Game where players can also decide to act sequentially. In this case, players in the same clique still seek assurance from each other, but a directed clique link from a clique x to a clique y now denotes that the players in clique x decide on whether or not to act, after having observed all players in clique y previously acting. Different in the PBEs of such a variant of the Assurance Game, is that if, for example, a leading clique z has two direct follower cliques y and x , it is not possible that clique v is a direct follower of clique w , but not of clique x . As a clique link now denotes the sequence in which players acts, clique v needs to be a direct follower of both cliques x and y . For the rest, it continues to be the case that players can lock each other into forming high endogenous thresholds. For example, in Figure 5(b), the five players in the leading clique simultaneously decide whether to act, and decide before the player in the one-player clique decides, because they believe that if they do not make their decision first, collective action never takes place, and because these beliefs are confirmed in equilibrium. Thus, a player with a low exogenous threshold may only decide whether or not to act after having observed the decision of a player with a high exogenous threshold.
 18. Indeed, Chwe (2000) focuses on examples characterized by homophily and a threshold-based social hierarchy.
 19. More generally, for a given maximal exogenous threshold t_{\max} , one could argue that as long as the number of players n is large enough, it is likely that there are no gaps in the exogenous threshold distribution. In this case, a bandwagon PBE assurance network as in Figure 6(a) always exists, where only threshold-2 players are in multi-player cliques, and all other players are in one-player cliques. Yet, as the examples show, players may still coordinate on forming multi-player cliques in this case.
 20. Note that while information⁺-proofness excludes from the set of all PBE assurance architectures, all architectures containing isolated one-player cliques, information⁻-proofness does not exclude any other architectures from this set. This is because the criterion to find an information⁻-proofness PBE assurance architecture is that we can find an Assurance Game that has an information⁻-proofness PBE assurance network with the given architecture. For any assurance network that does not have isolated one-player cliques, such a game is found by assigning to each node an exogenous threshold equal to the number of links formed, plus one—the same procedure that was used to find PBE assurance architectures more generally.
 21. Consider a PBE assurance network g that is not information^{+/-}-proof, and let players move to another PBE assurance network g' , in accordance with information^{+/-}-proofness. Then g' need not itself be information^{+/-}-proof,

- generating a further move to a new network g'' . It is easy to construct examples where, for example, the step from g to g' involves only deleting links, and the step from g' to g'' involves only adding links. In this interpretation, information^{+/−}-proofness still allows (step by step) for both adding and deleting links. Yet, Figure 8 illustrates that the consecutive adding and/or deleting of links need not lead to Pareto efficiency.
22. Pair-wise stability is applied to models with undirected links, where any link formed between players is automatically bilateral, and is only formed when both players agree to it. In our model, while a pair of players mutually forms links when it is in both their interests, to form a directed link to another player is still a unilateral decision by each player. Moreover, since in our model players do not only exchange information but also decide whether or not to act, and since two players can form a redundant mutual link by locking each other into only acting when both are in state w , a joint deviation may also involve deleting the links they have to each other.
 23. Jackson and Van den Nouweland (2005) introduce the concept of strong stability, where any set of players (two or more) may make any deviations (including jointly adding and deleting links).
 24. As Proposition 1 can be used to enumerate all PBE assurance architectures, aggregated across all Assurance Games we consider, and as Propositions A1 to A3 in Appendix 2 can be used to enumerate all Assurance Games with any given architecture, the combination of Proposition 1 and Propositions A1 and A3 allows one to enumerate all PBE (information^{+/−}-proof) assurance networks, aggregated across all Assurance Games.
 25. This corresponds to the correlated equilibrium that is often presented for a Battle of the Sexes game. Such a two-player two-strategy coordination game has two Nash equilibria, one of which is better to each player. Players may coordinate on a correlated equilibrium where they play one equilibrium on even days, and on the other equilibrium on odd days. In this manner, players' expected payoffs are equalized.
 26. Note that such an alternative model does *not* constitute a so-called repeated game, where the same game is played repeatedly by the same set of players. Each time, a new set of players plays the game, and in each separate game, players may have different characteristics.
 27. A third variant of the model with asymmetric information can be considered, where the individual player again does not know the exogenous thresholds of other players. When at Stage 1 player i forms a link to another player j , she does not only find out player j 's state, but also j 's exogenous threshold. At stage 2, having observed the exogenous thresholds and states of the players to which she formed a link, but not of the players to which she did not form a link, each player decides whether or not to act. Players who form few links need not run any risk. Simply, they may only act when the players to which links are formed have low exogenous thresholds (and are in state w). Yet, if linking costs are low, players will still form many links, in accordance with core-periphery networks, to ensure that collective action can also take place when other players have high exogenous thresholds.

28. Note that this includes the case where the set Z is empty, and where the set Y is identical to the set Z (so that in general, $I_z \geq I_y$).

References

- Aumann R (1959) Acceptable points in general cooperative n -person games. In: Tucker AW and Luce RD (eds) *Contributions to the Theory of Games IV*. Princeton, NJ: Princeton University Press, pp. 287–324.
- Aumann R and Brandenburger A (1995) Epistemic conditions for Nash equilibrium. *Econometrica* 63: 1161–1180.
- Bala V and Goyal S (2000) A noncooperative model of network formation. *Econometrica* 68: 1181–1229.
- Bang-Jensen J and Gutin G-Z (2009) *Digraphs: Theory, Algorithms and Applications*. London: Springer Verlag.
- Bernheim BD, Peleg B and Whinston MD (1987) Coalition-proof Nash equilibria I: concepts. *Journal of Economic Theory* 42: 1–12.
- Bikhchandani S, Hirshleifer D and Welch I (1992) A theory of fads, custom, and cultural change as informational cascades. *Journal of Political Economy* 100: 992–1026.
- Binmore K and Samuelson L (2001) Coordinated action in the electronic mail game. *Games and Economic Behavior* 35: 6–30.
- Bloch F and Jackson MO (2006) Definitions of equilibrium in network formation games. *International Journal of Game Theory* 34: 305–318.
- Bramoullé Y and Kranton R (2007) Public goods in networks. *Journal of Economic Theory* 135: 478–494.
- Brandenburger A (1996) Strategic and structural uncertainty in games. In: Zeckhauser R, Keeney R and Sebenius J (eds) *Wise Choices: Games, Decisions, and Negotiations*. Boston, MA: Harvard Business School Press, pp. 221–232.
- Carlsson H and Van Damme E (1993a) Equilibrium selection in stag hunt games. In: Binmore K, Kirman A and Tanki P (eds) *Frontiers of Game Theory*. Cambridge, MA: MIT Press, pp. 237–254.
- Carlsson H and Van Damme E (1993b) Global games and equilibrium selection. *Econometrica* 61: 989–1018.
- Centola DM (2013) Homophily, networks, and critical mass: Solving the start-up problem in large group collective action. *Rationality and Society* 25: 3–40.
- Chwe MS-Y (1999) Structure and strategy in collective action. *American Journal of Sociology* 105: 128–156.
- Chwe MS-Y (2000) Communication and coordination in social networks. *Review of Economic Studies* 67: 1–16.
- Corbae D and Duffy J (2008) Experiments with network formation. *Games and Economic Behavior* 64: 81–120.
- Fudenberg D and Tirole J (1991) Perfect Bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory* 53: 236–260.
- Goyal S and Vega-Redondo F (2005) Network formation and social coordination. *Games and Economic Behavior* 50: 178–207.

- Granovetter M (1978) Threshold models of collective behaviour. *American Journal of Sociology* 83: 1420–1443.
- Granovetter M (1985) Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91: 481–510.
- Güth W and Kliemt H (1994) Competition or co-operation: On the evolutionary economics of trust, exploitation and moral attitudes. *Metroeconomica* 45: 155–187.
- Güth W and Kliemt H (1998) The indirect evolutionary approach: Bridging the gap between rationality and adaptation. *Rationality and Society* 10: 377–399.
- Harsanyi JC and Selten R (1988) *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.
- Jackson MO (2010) *Social and Economic Networks*. Princeton, NJ: Princeton University Press.
- Jackson MO and Van den Nouweland A (2005) Strongly stable networks. *Games and Economic Behavior* 51: 420–444.
- Jackson MO and Watts A (2002) The evolution of social and economic networks. *Journal of Economic Theory* 106: 265–295.
- Jackson MO and Wolinsky A (1996) A strategic model of social and economic networks. *Journal of Economic Theory* 71: 44–74.
- Kim Y-G and Sobel J (1995) An evolutionary approach to pre-play communication. *Econometrica* 63: 1181–1193.
- Kreps DM and Wilson R (1982) Reputation and imperfect information. *Journal of Economic Theory* 27: 253–279.
- Kuran T (1988) The tenacious past: theories of personal and collective conservatism. *Journal of Economic Behavior and Organization* 10: 143–171.
- Kydd A (2000) Overcoming mistrust. *Rationality and Society* 12: 397–424.
- Lazer D (2001) The co-evolution of individual and network. *Journal of Mathematical Sociology* 25: 69–108.
- Lohmann S (1994) The dynamics of informational cascades. *World Politics* 47: 42–101.
- McPherson M, Smith-Lovin L and Cook JM (2001) Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27: 415–444.
- Macy MW (1991) Chains of cooperation: threshold effects in collective action. *American Sociological Review* 56: 730–747.
- Mengel F (2012) Learning across games. *Games and Economic Behavior* 74: 601–619.
- Oliver PE and Marwell G (1988) The paradox of group size in collective action: A theory of the critical mass II. *American Sociological Review* 53: 1–8.
- Oliver PE and Marwell G (2001) Whatever happened to critical mass theory? A retrospective and assessment. *Sociological Theory* 19: 292–311.
- Olson M (1965) *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.
- Robson A (1990) Efficiency in evolutionary games: Darwin, Nash and the secret handshake. *Journal of Theoretical Biology* 144: 379–396.
- Runge CF (1984) Institutions and the free rider: The assurance problem in collective action. *Journal of Politics* 4: 154–181.

- Sen A (1967) Isolation, assurance and the social rate of discount. *Quarterly Journal of Economics* 81: 112–124.
- Skyrms B (2004) *The Stag Hunt and the Evolution of Social Structure*. Cambridge, MA: Cambridge University Press.
- Spence M (1973) Job market signaling. *Quarterly Journal of Economics* 87: 355–374.
- Sunstein CR (2009) *Going to Extremes: How Like Minds Unite and Divide*. New York: Oxford University Press.
- Tucker A (1950) A two-person dilemma. In: Rasmussen E (ed.) *Readings in Games and Information*. Oxford: Blackwell, pp. 7–8.
- Weesie J, Snijders C and Buskens V (2009) The rationale of rationality. *Rationality and Society* 21: 249–277.
- Yin C-C (1998) Equilibria of collective action in different distributions of protest thresholds. *Public Choice* 97: 535–567.

Appendix I

Proofs of propositions in body of article

Proof of Lemma 1. In any candidate pure-strategy PBE, across all Assurance Games, short of obtaining M with certainty, the smallest risk that any player can face by acting is when having exogenous threshold 2, when she is willing, and when in equilibrium any other willing player automatically acts. In this case, the player still does not act if equation (1) is valid. Given equation (1), by the fact that $c > 0$, player i will only pay attention to other players' states if this allows her to find out that at least $(t_i - 1)$ other players are in state w .

For given strategies of the other players, consider the set of all strategies of player i with the following property: if all players of which player i checks the states are in state w , then she obtains M with certainty when acting. Consider in this set of strategies of player i with this property, the strategy with the weakly smallest number of links formed by player i , denoted τ_i . Then player i does not want to check the states of more than $(\tau_i - 1)$ other players if equation (3) is valid. The left-hand side of (3) takes on this value by the definition of τ_i , and because under the restriction of (1), player i only acts when all $(\tau_i - 1)$ checked other players are in state w . The right-hand side of (3) is constructed as follows. Suppose that player i checks x extra players. Then at best, any subset of at least $(\tau_i - 1)$ players who are all in state w suffices to eliminate for player i the possibility of incurring loss L when acting.

By the definition above of τ_i , as long as equation (1) is valid, if player i checks fewer than $(\tau_i - 1)$ players, she will not act. Checking a positive number of players smaller than $(\tau_i - 1)$ is therefore to no avail, so that

player i 's decision is between checking $(\tau_i - 1)$ other players and not checking any players. If player i is certain that each of the $(\tau_i - 1)$ players she checks, acts whenever being in state w , player i prefers to check $(\tau_i - 1)$ other players if equation (2) is valid.

For $M > (n - 1)c$, equation (2) is valid for any $\tau_i \leq n$ if $\varepsilon = 0$ and therefore also for a range of small ε . Furthermore, as equation (3) is valid for any $\tau_i \leq n$ if $\varepsilon = 0$, it is also valid for a range of small ε . Finally, for any positive ε for which equations (2) and (3) are valid, by equation (1), one can find large L such that equation (1) is valid.

Proof of Proposition 1

Step 1. In any PBE assurance network, each leading clique of active players has cardinality of at least 2.

Any player who does not form links to other players but still acts could only have threshold 1, which we have excluded in our analysis.

Step 2. In any PBE assurance network, if *any* player is in isolated one-player cliques, then there are at least two such players.

We prove this by contradiction. Consider a single player in an isolated one-player clique. Let n_1 players still form links to other players, meaning that these players act when finding out that all players they check are in state w . Then the unconnected player i could receive information that n_1 other players will act by connecting to the n_1 players. Therefore, the unconnected player will only not form links in equilibrium if $t_i > (n_1 + 1)$. At the same time, each of the connected players has exogenous threshold at most n_1 . It follows that if player i is the only player in an isolated one-player clique, it must be that $t_{\max} > (n_1 + 1)$, whereas the game has $n = (n_1 + 1)$ players, so that it must be that $t_{\max} > n$, which we have excluded.

Step 3. If a directed path $i \rightarrow j_1 \rightarrow j_2 \rightarrow j_3 \rightarrow j_4 \dots \rightarrow j_{m-1} \rightarrow j_m$ exists in a PBE assurance network, it must also be the case in this network that $i \rightarrow j_2$, $i \rightarrow j_3$, $i \rightarrow j_4$, and so on.

By assumptions (1)–(3), we have restricted the parameters in such a way that if a PBE assurance network contains link $i \rightarrow j_1$, player i only acts when being certain (in equilibrium) that player j_1 acts. But player j_1 by the same reasoning only acts when finding out that player j_2 is in state w . Given that information is local, it follows that player i will also form a link to player j_2 , and will only act when player j_2 is in state w . By the same reasoning, player

j_1 also forms a link to j_3 . Given this fact, i also forms a link to j_3 . In the same way, j_1 also forms a link to j_4 , which in turn means that i forms a link to j_4 , and so on. Given that any directed clique path contains a directed path of players, the claim extends to players located on any directed clique path.

Step 4. If a PBE assurance network contains a directed cycle, then all the players in this directed cycle are in one and the same clique.

This follows directly by applying Step 1 to all directed paths contained in the directed cycle. As any directed clique cycle also contains a directed cycle containing all players in the clique cycle, the claim follows.

Steps 5, 6, and 7 now use Steps 3 and 4 to establish that any PBE assurance network is a directed clique graph.

Step 5. Consider in a PBE assurance network any two cliques x and y such that there is at least one player i who is contained in both clique x and y . Then x and y are not maximal cliques, as all players in x and y are in one and the same clique in the PBE assurance network.

If there is a player i who is part of both cliques, then a directed cycle exists encompassing all players in these cliques, so that by Step 4, all these players need to be in one and the same clique. Note that given the possibility of one-player cliques, this shows that any PBE assurance network can be represented as a partition of maximal cliques with directed links between them.

Step 6. Suppose that in a PBE assurance network, player i in clique x forms a link to player j in clique y , then every player in clique x forms a link to every player in clique y .

By Lemma 1, player i only acts when being certain (in equilibrium) that player j will act, and player j only acts when finding out that every other player in clique y is in state w . It follows by Step 3 that player i only acts when finding out that every player in clique y is in state w .

Step 7. Consider in a PBE assurance network two cliques x and y such that there is no player who is contained in both cliques x and y . Then, if player i in clique x forms a link to player j in clique y , no player in clique y can form any link to a player in clique x .

If there is a player i in clique x who forms a link to a player j in clique y , and a player k in clique y who forms a link to a player l in clique x , then a

directed cycle exists containing all players in the two cliques, so that by Step 4, they must be in the same clique.

Step 8. Establishes that conditions (ii)–(iv), are also sufficient.

For any given acyclic directed clique graph with the properties (ii) to (iv), one can find an Assurance Game for which this is a PBE assurance network in the following way. First, for any player in the PBE assurance network, take the number of links the player forms (including the link to herself), and put her exogenous threshold equal to this number of links plus one. Second, given a number n_1 of players with links to other players, for any of the at least two isolated players, choose any threshold strictly larger than $(n_1 + 1)$ and smaller than or equal to n .

Proof of Proposition 2. Given that information^{+/-}-proof PBE assurance architectures are also PBE assurance architectures, they must meet all necessary conditions in Proposition 1, (i) to (iv). A PBE assurance network with isolated one-player cliques is not information^{+/-}-proof. Since $t_{\max} \leq n$, players in isolated one-player cliques will find it in their joint interest to connect to each other, and if necessary to a weak subset of the other players. It follows that condition (v) is added, stating that information^{+/-}-proof PBE assurance architectures do not contain isolated one-player cliques. To show that these conditions are also sufficient, the same procedure can be followed as in Proposition 1, namely, assign to each node in such a graph meeting conditions (i) to (v), an exogenous threshold equal to the number of links formed plus one. It is clear then that no single player, or subset of players, can do better by either deleting, or adding links.

Proof of Proposition 3. The existence of at least one PBE assurance network follows from the fact that by Proposition A3 below, the complete graph is a PBE assurance network for any Assurance Game. We show that each Assurance Game also has at least one information^{+/-}-proof PBE assurance network by induction.

We first show the inductive step. Let a set of n_1 players currently have links to other players, and assume that their strategies are in accordance with information^{+/-}-proofness (in the sense that if these were the only players, the graph consisting of their links would be an information^{+/-}-proof PBE assurance network). Consider the set of $n_2 = (n - n_1)$ players who currently do not have links, and rank them according to their exogenous thresholds, where the lowest exogenous threshold in this set is t_{\min} . Consider the cumulative frequency function of the number of players $n_2(t = t_{\min} + x)$ with exogenous threshold $(t_{\min} + x)$ or lower in this set (where $x = 0, 1, \dots, t_{\max}$).

Then given that $t_{\max} \leq n$, there must be at least one x such that $n_1 + n_2(t = t_{\min} + x) \leq t_{\min} + x$. It follows that, if the $n_2(t = t_{\min} + x)$ players all form links to one another as well as to a minimal subset of n'_1 of the n_1 players, the n_2 players form a number of links at least equal to their exogenous threshold minus 1. Consider next the smallest x in the cumulative frequency function such that $n_1 + n_2(t = t_{\min} + x) \leq t_{\min} + x$. Given the above, there must exist such a minimal x . Note that the clique determined by the smallest x meets all the conditions of Proposition A3 below. It follows that if the set of players would only include the n_1 players, plus the specified clique, the graph constructed would be an information^{+/-}-proof PBE assurance network.

We next derive the base case. This is simply obtained from the fact that the inductive step also applies to the case where $n_1 = 0$, that is, for the construction of a leading clique. Together, the base case and the inductive step show how an information^{+/-}-proof PBE assurance network can be constructed for any Assurance Game.

Proof of Proposition 4

- (i) For PBE assurance networks, the allowable exogenous thresholds of players in a leading clique with t_{\max} players, is a direct application of Proposition A3(ii). Given that $l_z = 0$, for any player with integer label $1 \leq q \leq s$, it is the case for her exogenous threshold t_q that $\max\{2, t_{q-1}\} \leq t_q \leq t_{\max}$. It follows that any player in the leading clique may have any exogenous threshold between 2 and t_{\max} . By Proposition A2(i), any player i in a one-player direct-follower clique may have any exogenous threshold $2 \leq t_i \leq (t_{\max} + 1)$. Yet, as no player's exogenous threshold can exceed t_{\max} , this rule becomes $2 \leq t_i \leq t_{\max}$.
- (ii) For information^{+/-}-proof PBE assurance networks, the allowable exogenous thresholds of players in a leading clique with t_{\max} players, is a direct application of Proposition A3(iii). Given that $l_z = 0$, for any player with integer label q such that $1 \leq q \leq s$, it is the case that $\min[\max\{1 + q, t_{q-1}\}, t_{\max}] \leq t_q \leq t_{\max}$. It follows that in the leading clique, there may never be more than $(x - 1)$ players with exogenous threshold x or lower (where $x \leq (t_{\max} - 1)$). The rule for the players in the one-player direct follower cliques is identical whether or not one considers information^{+/-}-proofness.

Proof of Proposition 5. By equation (3), a player who has the choice between, on the one hand, checking the states of $(t_{\max} - 1)$ or t_{\max} other players, and acting only when all of these players are in state w , and, on the other hand, not

acting, prefers to check. By equation (3), if checking $(t_{\max} - 1)$ or t_{\max} other players suffices to reach certainty that one's exogenous threshold of acting players is achieved, the individual player will not want to check more states.

We next show that each player needs to know that at least $(t_{\max} - 1)$ other players will act, before she is willing to act herself. Consider a player i with exogenous threshold 2 and in state w , and assume she has checked the states of $(t_{\max} - 2)$ other players. We show that for large L , such a player will still not want to act. Suppose that player i generously assumes that each checked player, when having exogenous threshold $(t_{\max} - 1)$ or lower, acts. Assume moreover that player i generously assumes that each player of which she does not check the state, acts when she is in state w . Then because player i has exogenous threshold 2, it suffices that at least one player has exogenous threshold $(t_{\max} - 1)$ or lower (put otherwise: it suffices that not all checked players have exogenous threshold t_{\max}), or that at least one non-checked player is in state w (put otherwise: not all non-checked players are in state u), for M to be obtained when acting. Player i still does not want to act if

$$\left[1 - f(t_{\max})^{t_{\max}-2} \varepsilon^{n-t_{\max}}\right] M + f(t_{\max})^{t_{\max}-2} \varepsilon^{n-t_{\max}} (-L) < 0$$

Finally, it is easy to see that a PBE assurance network containing a leading clique with more than t_{\max} players is not information-proof.

Appendix 2

Additional propositions

Proposition A1. In an acyclic directed clique graph with the properties given under Proposition 1, let n_1 out of n players form links to other players. Then in any PBE assurance network, for any player i in an isolated one-player clique, $(n_1 + 2) \leq t_i \leq n$. Furthermore, in any information^{+/−}-proof PBE assurance network, $n_1 = n$.

Proof. Consider a player i in an isolated one-player clique. Let n_1 players still form links to other players. Then player i can potentially find out that $(n_1 + 1)$ will act by connecting to the n_1 players. Therefore, player i will only *not* form links if $t_i > (n_1 + 1)$. At the same time, t_i should not exceed n , simply because t_{\max} cannot exceed n .

For any PBE assurance network g with $(n - n_1)$ players in isolated one-player cliques, a PBE assurance network g' exists where these $(n - n_1)$ players together form a single clique, which forms clique links to *all* cliques that were already contained in g . Simply, if all other players in the clique of

$(n - n_1)$ players only act when finding out that all other players in the clique are in state w , as well as all the n_1 players in g , then it is a best response for a single of the $(n - n_1)$ players to form all these links. As the obtained PBE assurance network leaves each of the $(n - n_1)$ players better off, no PBE assurance network with isolated one-player cliques is an information^{+/-}-proof PBE assurance network.

Proposition A2. In any acyclic directed clique graph with properties (ii) to (iv) in Proposition 1, consider any non-isolated one-player clique x . Denote by l_y the total number of links that player i in clique x forms to the set of cliques Y (so that $(l_y + 1)$ is i 's endogenous threshold). Denote by l_z the largest number of links *strictly* smaller than l_y that player i can obtain by connecting to a set of cliques Z different from Y . Then in any Assurance Game for which the acyclic directed graph is an (information^{+/-}-proof) PBE assurance network, player i 's exogenous threshold t_i :

- (i) may take on any value such that $(l_z + 2) \leq t_i \leq (l_y + 1)$;
- (ii) may be smaller than or equal to the endogenous threshold τ_φ of a player j to which player i forms a link, if player j is in the unique clique y which is on a longest directed clique path of length 1 from player i , where $t_i = \tau_j$ is only allowed if this clique y contains at least two players, and $t_i < \tau_j$ is only allowed if clique y contains at least three players.

Proof

- (i) If $t_i < (l_z + 2)$, player i can do better by forming links to the set of cliques Z instead. If $t_i > (l_y + 1)$, under the assumptions in equations (1)–(3), player i cannot form a sufficient number of links to eliminate all risk from acting in equilibrium. For any $(l_z + 2) \leq t_i \leq (l_y + 1)$, the (weakly) best that player i can do is to continue to hold her current links, even though this may mean that $t_i < \tau_j$.
- (ii) It cannot be that $t_i < \tau_j$ for player j in a one-player clique y to which i forms a link, as otherwise player i prefers to remove the link to j . Furthermore, it cannot be that $t_i < \tau_j$ if clique y lies on a longest directed clique path from x with length larger than 1, because otherwise player i would not want to link to the cliques between x and y on this directed clique path. There cannot be another clique z on a longest directed clique path with length 1 from x , because player i is already obtaining information in excess of her exogenous threshold from clique y , and since clique y by definition does not form a link to clique z . If player i is in a clique x on a longest directed clique

path of length 1 from a clique y with exactly two players, then $t_i < \tau_j$ is impossible, because player i can get information about $(\tau_j - 2)$ players by not connecting to clique y (i.e. the $(\tau_j - 2)$ links that players in clique y form to players outside of their clique), which for $t_i \leq (\tau_j - 1)$ player i would prefer to do. Yet, it can now still be the case that $t_i = \tau_j$, as the information obtained from $(\tau_j - 2)$ players by no longer connecting to clique y does not suffice then. If player i is in a clique x on a longest directed clique path of length 1 from a clique y with n_y players ($n_y \geq 3$), then it may be that $t_i < \tau_j$ or $t_i = \tau_j$, because player i can get information about $(\tau_j - n_y)$ players by not connecting to clique y (i.e. the $(\tau_j - n_y)$ links that players in clique y form to players outside of their clique); for $n_y \geq 3$, it is possible then that $(\tau_j - n_y + 1) < t_i \leq \tau_j$.

Proposition A3

In any acyclic directed graph g with the properties (ii) to (iv) in Proposition 1, consider any clique x with $s \geq 2$ players, and label these players $1, 2, \dots, p, q, r, \dots, s$, such that $t_p \leq t_q \leq t_r$. Denote by l_y the total number of links that any player q in clique x forms to players *not* in clique x , in a set of cliques Y (so that the endogenous threshold of player q equals $(l_y + s)$), and denote by l_z the largest number of links *strictly* smaller than $(l_y + s - 1)$ that any player q can obtain by no longer connecting to any other player in clique x , and connecting to all players in a set of cliques Z .²⁸ The following now applies:

- (i) If g is a PBE assurance network, then for any $i \rightarrow j \in g$ such that $i \in x$ and $j \notin x$, $t_i > t_j$;
- (ii) If g is a PBE assurance network, then for any player with integer label q such that $1 \leq q \leq s$, it is the case that $\max\{l_z + 2, t_{q-1}\} \leq t_q \leq (l_y + s)$ (where $t_0 = 0$);
- (iii) If g is an information^{+/−}-proof PBE assurance network, then for any player with integer label q such that $1 \leq q \leq s$, it is the case that $\min\left[\max\{l_z + 1 + q, t_{q-1}\}, (l_y + s)\right] \leq t_q \leq (l_y + s)$ (where $t_0 = 0$).

Proof

- (i) Given that x has at least two players, if the exogenous threshold of a player j outside of clique x to which player i in clique x forms a link is such that $t_i \leq t_j$, player i finds it better to stop linking to the players in clique x .
- (ii) If $t_q \leq (l_z + 1)$, then player q in clique x is better off by linking to all players in the defined set of cliques Z . t_q by definition may also be

constrained by the exogenous threshold assigned to the player with a threshold just lower, because of our convention of giving players in clique x with weakly higher exogenous thresholds a higher label. If $t_q > (l_y + s)$, then the graph cannot be a PBE assurance network by Lemma 1. For any other levels of t_q , player q cannot do better by reconnecting, and at the same time in equilibrium can obtain enough information to be able to act without risk.

- (iii) For player 1 in clique x , if $(l_z + 2) = (l_y + s)$, then by the way in which players are labeled, every player q should have $t_q = (l_y + s)$. If instead $(l_z + 2) < (l_y + s)$, given the convention that $t_0 = 0$, the condition says that $t_1 \geq (l_z + 2)$. If not, player 1 would be better off by connecting to set of cliques Z . If $t_1 = (l_z + 2) < (l_y + s)$, then the condition says that player 2 should have an exogenous threshold at least $t_2 \geq (l_z + 2)$; note that if this is not valid, there are two players with exogenous threshold $(l_z + 2)$ who can coordinate on doing with fewer links; if $t_1 > (l_z + 2)$, then given that players are labeled such that players with a weakly higher exogenous threshold have a higher label, player 2 should have an exogenous threshold at least . In general, unless constrained by the exogenous threshold of the player with the label just below, and as long as $(l_z + 1 + q)$ does not lie above $(l_y + s)$, player q should have an exogenous threshold at least $(l_z + 1 + q)$. If not, there would be q players with exogenous threshold $(l_z + q)$ or lower who can all do better by no longer connecting to the other players in their clique.