Essays on bargaining and strategic communication

de Groot Ruiz, A.W.

Publication date
2012

Chapter 3 ACDC Rocks When Other Criteria Remain Silent[21]

3.1 Introduction

Crawford & Sobel (1982) showed how meaningful costless communication between an informed Sender and an uninformed Receiver can be supported in equilibrium. Their seminal paper inspired many applications ranging from the presidential veto (Matthews, 1989), legislative committees (Gilligan & Krehbiel, 1990) and political correctness (Morris, 2001) to double auctions (Matthews & Postlewaite (1989); Farrell & Gibbons (1989)), stock recommendations (Morgan & Stocken, 2003) and matching markets (Coles, Kushnir & Niederle, 2010). These cheap talk games are characterized by multiple equilibria which differ crucially in their prediction about how much information will be transmitted. Standard signaling refinements such as Kohlberg & Merten's strategic stability (1986) have no bite in this setting, as messages are costless. Hence, this raises the important issue of equilibrium selection. As of yet, however, no satisfying refinement exists that works well across a wide range of cheap talk games.[22]

To overcome the selection problem, Farrell (1993) pioneered the approach of what we call credible deviations. Farrell observed that, in contrast to what is assumed in standard game theory, communication in real life is based on a pre-existing natural language.[23] Hence, it is possible to send out-of-equilibrium messages that will be understood, although they will not necessarily be believed. Farrell proposes a criterion under which such out-of-equilibrium messages, or neologisms, are credible and, hence, should be believed by a Receiver. Arguably, an equilibrium is not stable if some Senders can send a credible neologism that would induce a rational Receiver to deviate from equilibrium. Equilibria that do

[21] This chapter is based on De Groot Ruiz, Offerman & Onderstal (2011b).
[22] For a comprehensive review of Sender-Receiver games, see Sobel (2010).
[23] See Blume, DeJong, Kim & Sprinkle (1998) and Agranov & Schotter (2010) for studies on the role of natural language in cheap talk games.

not admit credible neologisms are ‘neologism proof.’ Alas, neologism proofness tends to be too effective and regularly eliminates all equilibria. Matthews, Okuno-Fujiwara & Postlewaite (1991) address important conceptual issues with neologism proofness and propose three alternative accounts of what constitutes a credible deviation. Unfortunately, these ‘announcement proofness’ criteria also fail to be predictive in many games (such as the Crawford-Sobel (1982) model). Partly for this reason, several other types of concepts have been proposed that distinguish between stable and unstable equilibria (or profiles), such as Partial Common Interest (PCI) (Blume, Kim & Sobel, 1993), the recurrent mop (Rabin & Sobel, 1996) and No Incentive To Separate (NITS) (Chen, Kartik & Sobel, 2008). These criteria often work well in specific settings, but fail to discriminate successfully across across a wider range of cheap talk games.

We take a different tack: we propose a criterion that is based on credible deviations but allows for a continuous instead of a binary stability concept. Our first conjecture is that credible deviations matter, in the sense that an equilibrium which does not admit credible deviations is more stable than one that does. Our second conjecture is that credible deviations matter gradually: of two equilibria that allow for credible deviations, the one with smaller deviations will predict better. A binary stability criterion is appropriate for rational agents, but may unnecessarily lose predictive power when applied to human behavior, which is seldom completely in (or out of) equilibrium.

We formalize this idea in the Average Credible Deviation Criterion (ACDC). According to ACDC, the behavioral stability of an equilibrium is a decreasing function of its Average Credible Deviation (ACD), a measure of the frequency and intensity of credible deviations. The ACD measures the mass of types that can credibly deviate and the size of those induced deviations (as measured by the difference in Sender payoff between the equilibrium and deviating action). Comparable equilibria will perform better if they have a lower ACD on this account. In particular, we call an equilibrium that minimizes the ACD in a game an ‘ACDC equilibrium.’ This allows us to select equilibria, even in games where no equilibrium is completely stable.

We show that an ACDC equilibrium exists under general conditions. In addition, ACDC makes meaningful predictions in a variety of settings that have previously been analyzed theoretically and experimentally. We look at how ACDC performs in a range of discrete games analyzed by Blume, DeJong, Kim & Sprinkle (2001) in the laboratory and find that ACDC organizes the main features of the data. We illustrate the predictive capability of ACDC in a veto threats model with a large equilibrium set where existing criteria fail. This game belongs to the class of veto threats games introduced in Chapter 2 and we show here that ACDC selects a unique most informative equilibrium in this class. In Chapter 4, we study this class of games experimentally and find that the results corroborate the predictions of ACDC. Furthermore, we show that ACDC selects the unique maximum size equilibrium in the leading uniform quadratic case of the Crawford-Sobel game for a large range of bias parameters. Until now, only NITS (Chen, Kartik & Sobel, 2008) was able to select (this equilibrium) in the Crawford-Sobel setting. In addition, the maximum size equilibrium becomes more stable as the bias parameter becomes smaller according to ACDC, which is not predicted by existing criteria. Both results are qualitatively supported by experimental work on (discrete) Crawford-Sobel games (Dickhaut, McCabe & Mukherji (1995), Cai & Wang (2006) and Wang, Spezio & Camerer (2010)).

ACDC is meant to solve a practical problem: how to select the most plausible equilibrium in cheap talk games? It constitutes an intuitive, generally applicable and experimentally validated selection criterion. In addition to selecting the most plausible equilibrium, ACDC can be informative about how well this equilibrium will perform if it is not completely stable. A major advantage is that it can easily be applied in almost all cheap talk games and provides meaningful results. Wherever experimental evidence exists, the predictions of ACDC are in line with the data: it performs at least as well as other criteria, when they are predictive, and also makes predictions when other criteria are silent. We believe these characteristics make it a valuable contribution to the current literature.

This chapter has the following structure. In section 3.2, we motivate and introduce ACDC and compare it to neologism proofness and announcement proofness. In section 3.3, we illustrate how ACDC works in some simple discrete games. In section 3.4, we compare ACDC to other concepts in a veto threats game with a large equilibrium set. In section 3.5, we analyze the ACDC-properties of the uniform quadratic Crawford-Sobel model. In section 3.6, we look at the performance of ACDC in experiments. Finally, section 3.7 concludes. All proofs are relegated to Appendix 3.8.

3.2 ACDC

ACDC is based on theories of credible deviations. To illustrate the idea behind credible deviations, we start with Game A, a simple game taken from Farrell (1993). A Sender sends a costless message m to the Receiver, who then takes one of three actions: a_1, a_2 or a_3. Both players' payoff functions depend on the Receiver's action and the Sender's type. The Sender's type, drawn from $\{t_1, t_2\}$ with equal probability, is private information of the Sender. Table 3.1 shows the payoffs.

TABLE 3.1
GAME A

	a_1	a_2	a_3
$t_1\,(\frac{1}{2})$	3, 3	0, 0	2, 2
$t_2\,(\frac{1}{2})$	0, 0	3, 3	2, 2

Notes: The left column shows the Sender's type and between brackets the probability that it is drawn. The top row shows the Receiver's actions. The remaining cells provide the Sender's payoff in the first entry and the Receiver's payoffs in the second entry.

The game has two (perfect Bayesian) equilibria: a pooling and a separating equilibrium.[24] In the pooling equilibrium, all Senders use the same message strategy and the Receiver always proposes a_3, regardless of the message. In the separating equilibrium, t_1 sends a message inducing a_1 whereas t_2 sends a different message that induces a_2.[25] The separating equilibrium seems more plausible, as it Pareto-dominates the pooling equilibrium. Still, standard signaling refinements in the vein of Kohlberg and Mertens' (1986) strategic stability

[24] An equilibrium outcome specifies which action(s) are played by the Receiver given each type. Each equilibrium outcome is induced by a whole class of essentially equivalent equilibria, which only differ in the messages used. For simplicity, we will refer to an equilibrium outcome simply by 'equilibrium.'

[25] We say a type t induces an action a, if she always sends a message to which the Receiver always responds with action a.

are silent in cheap talk games, since all messages are equally costless. For a similar reason, the (Agent) Quantal Response Equilibrium (McKelvey & Palfrey, 1998), which can often select equilibria in signaling games, is not predictive in cheap talk games.[26]

To overcome the selection problem, Farrell (1993) pioneered the idea of credible deviations and introduced the neologism proofness criterion. Neologisms are out-of-equilibrium messages which are assumed to have a literal meaning in a pre-existing natural language. Farrell considers neologisms which literally say: "play action $\tilde{a}$, because my type is in set N." Farrell deems a neologism credible if and only if (*i*) all types t in N prefer $\tilde{a}$ to their equilibrium action $a^{\sigma}(t)$, (*ii*) all types t not in N prefer their equilibrium action $a^{\sigma}(t)$ to $\tilde{a}$ and (*iii*) the best reply of the Receiver after restricting the support of his prior to N is to play $\tilde{a}$. We will denote neologisms by $\langle \tilde{a}, N \rangle$. According to Farrell, credible deviations lead rational players to deviate from equilibrium. An equilibrium is neologism proof, and stable on this account, if and only if it does not admit any credible neologism.

Matthews, Okuno-Fujiwara & Postlewaite (1991) identify two potential problems with this credibility criterion. First, it can sometimes be too strict, as it does not allow different deviating types to separate. Second, it is sometimes not strict enough. The credibility of a neologism is not affected by the existence of other credible neologisms, although the choice of the Sender about which neologism she sends potentially reveals information. To solve these problems they consider more elaborate credible deviations, called announcements. Like a neologism, an announcement specifies a type-subset N to which the deviating Sender claims to belong. An announcement also provides a strategy specifying for each types if and how they deviate. Let $N(t)$ be the set N type t claims to be in and $\mathcal{N}$ the set of all $N(t)$. Let β_0 be the Receiver's prior beliefs and β_0^N

[26] The Agent Quantal Response Equilibrium (A-QRE) is the extensive form game variant of the Quantal Response Equilibrium. The problem in cheap talk games is that the A-QRE does not eliminate the pooling equilibrium, which is often implausible. The pooling equilibrium is always a limiting, principal branch A-QRE: For any rationality parameter λ, there is an A-QRE where all Senders mix uniformly over the message space and the Receiver ignores all messages. As λ increases, the Senders strategy remains unchanged, and the Receiver's best response smoothly approaches its actual best response to her prior.

her prior beliefs β_0 restricted to have support N. Finally, let $BR(\beta)$ be the Receiver's best response set if she has beliefs β. An announcement is credible if (*i*) each deviating type prefers any action in $BR\left(\beta_0^{N(t)}\right)$ to her equilibrium action, (*ii*) each deviating type prefers any action in $BR\left(\beta_0^{N(t)}\right)$ to any action in $\bigcup_{N \in \mathcal{N} \setminus \{N(t)\}} BR\left(\beta_0^N\right)$, (*iii*) each non-deviating type prefers her equilibrium action to any action in $\bigcup_{N \in \mathcal{N}} BR\left(\beta_0^N\right)$ and (*iv*) for each deviating type no announcements exist satisfying (*i*)-(*iii*) where she earns more than in the current announcement.[27] They also define a weaker and a stronger credibility criterion. A weakly credible announcement only satisfies (*i*)-(*iii*), whereas a strongly credible announcement also must propose an equilibrium. Matthews and coauthors express a preference for (ordinary) credible announcements.[28]

In game A, the pooling equilibrium admits two credible neologisms, $\langle a_1, \{t_1\} \rangle$ and $\langle a_2, \{t_2\} \rangle$. These deviations also form a credible announcement. Hence, the pooling equilibrium is neither neologism proof nor announcement proof. The separating equilibrium, on the other hand, must be neologism and announcement proof since all types receive their maximum payoff. Hence, neologism and announcement proofness provide a compelling strategic reason why rational players would play the separating equilibrium.

Unfortunately, neologism and announcement proofness are often too effective and eliminate all equilibria, as Game B in Table 3.2 illustrates. Now three types can be drawn, t_1 and t_2 each with probability $(1-\delta)/2$ and t_3 with probability δ. The Receiver's best response is to play a_i if he knows the Sender is t_i. His best response is to play a_5 if he holds his prior beliefs and to play a_4 if he restricts his prior to have support $\{t_2, t_3\}$ (for $\delta < \frac{1}{2}$). t_1 prefers a_5 over all other actions and hence would prefer to pool with the other types. t_3 prefers to be identified as herself, as she prefers a_3. t_2 prefers to pool with t_3 (as she prefers

[27] We use a somewhat simpler definition than Matthews et al. for ease of exposition.

[28] Another approach similar to credible neologisms is Myerson's (1989) credible negotiation statements. Myerson is able to obtain a solution concept that guarantees existence but at the cost of assuming the presence of a mediator.

a_4) and the worst case for her is to be identified. The game has two equilibria. In the pooling equilibrium, all Senders induce a_5. In the semi-separating equilibrium, t_1 induces a_1, whereas t_2 and t_3 induce a_4. Neither equilibrium is neologism or announcement proof (if $0 < \varepsilon < 1$). The pooling equilibrium admits the credible neologism $\langle a_4, \{t_2, t_3\} \rangle$. The semi-separating equilibrium admits the credible neologism $\langle a_3, \{t_3\} \rangle$. In this game, (weakly) credible announcements coincide with credible neologisms.[29]

TABLE 3.2
GAME B

	a_1	a_2	a_3	a_4	a_5
t_1 $(\frac{1-\delta}{2})$	1, 4	0, 0	0, 0	0, 0	2, 3
t_2 $(\frac{1-\delta}{2})$	0, 0	0, 2 + δ	1 + ε, 0	2, 2	1, 1
t_3 (δ)	0, 0	0, 0	2 + ε, 3	2, 2	1, 1

Notes: We use the same notation as in Table 3.1. $0 < \delta < \frac{1}{2}$ and $0 < \varepsilon < 1$

For entirely rational agents, the fact that neither equilibrium is stable might be all there is to be said. When explaining or predicting human behavior, however, we feel we can go further. Human behavior is hardly ever completely in or out of equilibrium, and by imposing a binary distinction between stable and unstable equilibria one may lose predictive power.[30]

In game B, even though the separating equilibrium is not entirely stable, it seems more plausible than the pooling equilibrium if either t_3 is infrequent (δ small) or t_3 has a very small incentive to deviate (ε small). If δ is small, then the separating equilibrium will be upset with a small probability, whereas the pooling equilibrium will be upset almost half of the time. Similarly, if ε is small,

[29] Neither equilibrium admits strongly credible announcements, since the deviations profile does not comprise an equilibrium in itself. This is typical for many cheap talk games. Due to the strong credibility requirement of strongly credible announcements, in most games no equilibrium admits them.

[30] We consider equilibrium, when applied to human behavior, to be most meaningful in a dynamic context, where members of a group interact frequently with different other members. In this context language evolves and behavior is shaped by strategic forces in the direction of equilibrium. For a one-shot game between rational individuals without social information, an approach based on rationalizability and some focal meaning of messages, such as that in Rabin (1990), may be appropriate.

then t_3 has a small incentive to deviate in the separating equilibrium and may choose to stick to it, lest she be misunderstood and get a payoff lower than she gets by sticking to equilibrium. Hence, we would expect to observe behavior close to the separating equilibrium more frequently than behavior close to the pooling equilibrium.

Our intuition is that the behavioral stability of an equilibrium is a decreasing function of the average *intensity* of the credible deviations it admits. This depends, firstly, on the mass of types that can credibly induce a deviation and, secondly, on the intensity of the deviation, measured by the incentive the Sender has to deviate. As a consequence, if the deviating mass and the induced deviations from equilibrium are small, the equilibrium is likely to be a good predictor of behavior. We formalize this intuition in our ACDC criterion.

We will now define ACDC for a general setting that allows for a discrete and continuous type set, and covers both pure advising and veto threats bargaining games. We consider the following two-player cheap talk game. The game is played by a Sender and a Receiver. Nature draws the Sender type t from distribution f on T, where T is a compact metric space. The Sender then privately observes her type t and chooses a message $m \in M$. After having observed the Sender's message, the Receiver chooses an action $a \in A$, where A is a compact metric space.[31] After seeing the action, the Sender chooses between accepting ($v = 1$) or rejecting ($v = 0$) the action. If she rejects, the outcome is the disagreement point δ. If $\delta \in A$, the game has an interior veto threat and otherwise it has an exterior veto threat. In cheap talk games without veto threats, the Sender is forced to accept the action so that v is always equal to 1. The outcome set is $X = A \cup \{\delta\}$. Let $U^R : X \times T \to \mathbb{R}$ be the utility function of the Receiver $U^S : X \times T \to \mathbb{R}$ that of the Sender. We assume both are bounded from above and below. For a set S, let ΔS denote the set of probability distributions on S. Then, a strategy for the Sender consists of a function $\mu : T \longrightarrow \Delta M$ and a function $\nu : A \times T \longrightarrow [0,1]$ and a strategy of the Receiver is a function

[31] This representation allows for T and A to be de facto discrete, by having $U^S(t,x)$ and $U^R(t,x)$ be constant on regions of the type and outcome space.

$\alpha : M \longrightarrow \Delta A$. Let Σ^S be the set of Sender strategies and Σ^R the set of Receiver strategies. Let $\sigma \equiv \{\mu, \nu, \alpha\}$ be a strategy profile and Σ the set of all strategy profiles. $m(t)$ will denote the random variable of the Sender's message after learning t (determined by μ) and $a(m)$ will denote the random variable of the Receiver's action after receiving m (determined by α). Finally, let the Receiver have prior and posterior beliefs (a function $\beta : M \longrightarrow \Delta T$) over the Sender's type, $\beta^0 = f$ and $\beta(m)$ respectively.

The Sender's expected payoff is $EU^S(t, \mu, \nu \mid \alpha) = \int_{\Sigma} \left(\nu(a(m(t)), t) \cdot U^S(a(m(t)), t) + \left(1 - \nu(a(m(t)), t)\right) \cdot U^S(\delta, t)\right) d\sigma$. The Receiver´s expected payoff given her beliefs and message m is $EU^R(\alpha, m \mid \nu, \beta) = \int_{\Sigma^R} \int_{[0,1]} \int_T \left(\nu(a(m(t)), t) \cdot U^R(a(m), t) + \left(1 - \nu(a(m(t)), t)\right) \cdot U^R(\delta, t)\right) \left(d\beta(m)\right) d\nu d\alpha$. A perfect Bayesian equilibrium $\sigma^* = \{\{\mu^*, \nu^*, \alpha^*\}, \beta^*\}$ specifies a strategy profile together with Receiver beliefs and is characterized by the following three conditions:

$$\{\mu^*, v^*\} \in \arg\max_{\{\mu, v\} \in \Sigma^S} EU^S(t, \mu, \nu \mid \alpha^*) \text{ for all } t \in T$$

$$\alpha^* \in \arg\max_{\alpha \in \Sigma^R} EU^R(\alpha, m \mid, \nu^*, \beta^*) \text{ for all } m \in M \tag{3.1}$$

$$\beta^*(m) \text{ is derived from } \mu \text{ and } \beta^0 \text{ using Bayes Rule whenever possible}$$

Let Σ^* be the set of perfect Bayesian equilibria (from now on just 'equilibria'). ACDC provides a stability measure and a selection criterion for equilibria. The starting point of ACDC is a theory of credible deviations γ. Such a theory associates a deviating profile $\gamma(\sigma^*) \in \Sigma$ with an equilibrium σ^*. (For convenience, we write $\gamma(\sigma^*) = \{\mu^{\gamma*}, \nu^{\gamma*}, \alpha^{\gamma*}\}$) A deviating profile specifies firstly which Sender types would deviate and in which way, and secondly, how the Receiver would react. If no type can send a credible deviation according to γ, then $\gamma(\sigma^*) = \sigma^*$.

The following step is to measure the *intensity* of each type's credible deviation, which we denote by $CD_\gamma(t,\sigma^*)$. We would like this measure to have some properties. It should be

- invariant to affine transformations of payoffs;
- increasing in the difference between the deviating and equilibrium payoff;
- 0 if the difference between deviating and equilibrium payoff is zero;
- 1 if the difference between deviating and equilibrium payoff is maximal.

We then define the Average Credible Deviation (ACD) of an equilibrium σ^* relative to γ as:

$$ACD_\gamma(\sigma^*) = E_t[CD_\gamma(t,\sigma^*)] \tag{3.2}$$

Specifications of $CD_\gamma(t,\sigma^*)$ adhering to the properties above will lead to similar conclusions. We propose the following function for $CD_\gamma(t,\sigma^*)$. Let $\Sigma^\dagger$ be the set of rationalizable strategy profiles. Then, $\underline{U}^S(t) \equiv \inf_{\sigma\in\Sigma^\dagger} EU^S(t,\mu,v \mid \alpha)$ and $\bar{U}^S(t) \equiv \sup_{\sigma\in\Sigma^\dagger} EU^S(t,\mu,v \mid \alpha)$ are the lowest and highest rationalizable payoff for Sender type t. We define the credible deviation of type t in equilibrium σ^* relative to γ as

$$CD_\gamma(t,\sigma^*) = \frac{EU^S(t,\mu^{\gamma*},v^{\gamma*} \mid \alpha^{\gamma*}) - EU^S(t,\mu^*,v^* \mid \alpha^*)}{\bar{U}^S(t) - \underline{U}^S(t)} \tag{3.3}$$

if $EU^S(t,\mu^*,v^* \mid \alpha^*) > \underline{U}^S(t)$. If $EU^S(t,\mu^*,v^* \mid \alpha^*) = \underline{U}^S(t)$, then $CD_\gamma(t,\sigma^*) = 1$, as in this case the Sender has no incentive to adhere to her equilibrium strategy.

A deviation theory can be based on credible neologisms or credible announcements (or, in principle, on any theory of credible deviations). In the case that types can send multiple credible deviations, a deviating theory should also specify which one(s) Senders would use.[32] In many games, credible neologisms

[32] In the case of credible neologisms, a reasonable specification is that types coordinate on the credible neologism with the highest total intensity (mass-weighted sum of credible deviations). Results will in most games be qualitatively the same for different coordination rules.

and credible announcements coincide. Which theory predicts behavior best when they do not coincide, is an empirical question about which actually very little is known. We prefer credible announcements for 'simple' games and credible neologisms for 'complex' games. Credible neologisms have the virtue of simplicity, whereas credible announcements meet the soundest criterion of credibility. In simple games, human agents will be able to reason according to the logic of credible announcements and credible announcements seem most appropriate. In addition, a deviation theory based on credible announcements has the elegant property that it need not specify which credible announcement a type will play if she can send more than one, as they all must yield an identical payoff for each deviating Sender. In complex games, however, the complexity of announcements and the strict credibility criteria seem too demanding for boundedly rational agents. This can lead to counterintuitive predictions. For example, in the Crawford-Sobel game, announcement proofness may only select the pooling equilibrium, which seems an unlikely outcome for people to play.[33] Hence, here the simplicity of credible neologisms seems more appropriate. A rule of the thumb is to use credible announcements in discrete games and credible neologisms in continuous games.[34] We adhere to this rule of thumb in this chapter and, hence, when we apply ACDC we will suppress γ.

Based on the ACD, we formulate the ACD-Criterion (ACDC), which says that an equilibrium σ^* will on average predict better than equilibrium σ if $ACD_\gamma(\sigma^*) < ACD_\gamma(\sigma)$. In particular, based on ACDC we can formulate the following selection criterion:

[33] For instance, in the uniform quadratic game for $b \in \left(\frac{1}{24}, \frac{1}{16}\right)$, the pooling equilibrium is announcement proof, while the size-2 and size-3 equilibria are not. For these values of b, the pooling equilibrium admits the weakly credible announcement composed of the neologisms at the beginning and end, characterized by the set of intervals of deviating types sending the same message $\left\{\left[0, \frac{1}{6} - \frac{4}{6}b\right], \left[\frac{5}{6} - \frac{4}{6}b, 1\right]\right\}$. In addition, however, it admits the weakly credible announcement $\left\{\left[0, \frac{1}{5} - \frac{16}{5}b\right], \left[\frac{1}{5} - \frac{16}{5}b, \frac{2}{5} - \frac{12}{5}b\right], \left[\frac{3}{5} - \frac{12}{5}b, \frac{4}{5} - \frac{16}{5}b\right], \left[\frac{4}{5} - \frac{16}{5}b, 1\right]\right\}$. Since for all weakly credible announcements deviating types exist that prefer another weakly credible announcement, none is a credible announcement. The size-2 and size-3 equilibria only admit weakly credible announcements composed of the non-overlapping credible neologisms, which are thus credible announcements. Observe that the computational demands on agents to determine whether credible announcements exist and how they look like are quite high in this game.

[34] Admittedly, this rule is somewhat coarse, but does prevent ex-post determination as to what constitutes a 'simple' game.

Definition 3.1 *An equilibrium σ^* is an ACDC equilibrium relative to deviation theory γ if $ACD_\gamma(\sigma^*) \leq ACD_\gamma(\sigma)$ for all $\sigma \in \Sigma^*$.*

Note that this selection criterion selects the equilibrium that will predict best on average rather than the equilibrium that will always be played.

A simple implication is that σ^* is an ACDC equilibrium if $\gamma(\sigma^*) = \sigma^*$. Hence, the prediction of ACDC coincides with that of the underlying deviation theory if the latter identifies a stable equilibrium. The following result is immediate.

Proposition 3.1 *If the number of equilibrium outcomes is finite, the cheap talk game has an ACDC equilibrium relative to γ.*

Hence, existence of an ACDC equilibrium is guaranteed by a finite set of equilibrium-outcomes. This is a relevant result, as Park (1997) has shown that finite Sender-Receiver games have a finite set of equilibrium outcomes under generic conditions. Before, Crawford and Sobel (1982) showed a similar result for their setting with a continuous type-space.

Even when games do not have a finite outcome set (as our continuous veto threats model in section 3.4), mild conditions can be formulated in order to guarantee existence of an ACDC equilibrium:

Proposition 3.2 *Let s be an equilibrium outcome and $ACD_\gamma(s)$ the ACD of equilibria inducing s. Suppose the equilibrium outcome set S can be represented by a finite union of compact metric spaces $S = \bigcup_{i \in N} S_i$, such that $ACD_\gamma(s)$* is *continuous in s on all subsets S_i. Then, an ACDC equilibrium exists with respect to γ.*

For instance, this means that continuous games with an equilibrium set consisting of partition equilibria that are well-behaved with respect to their ACD will have an ACDC equilibrium.[35] In almost all applications we have come across, a unique ACDC equilibrium outcome exists. Still, one can construct games with multiple ACDC equilibrium outcomes (for instance when multiple outcomes are completely stable). In such cases, we are content with the conclusion that several equilibrium outcomes are equally plausible.

3.3 Discrete games

To illustrate ACDC, let us first put it to work in Game B (Table 3.2). In the pooling equilibrium, all types induce a_5. It admits the credible announcement where types t_2 and t_3 deviate to a_4. (Credible announcements and credible neologisms coincide in all games presented in this section.) In a discrete game, the ACD of a pure equilibrium σ (with a pure $\gamma(\sigma)$) reduces to

$$(3.4) \qquad \sum_{t\in T} f(t)\frac{U^S(a^{\gamma(\sigma)}(t),t)-U^S(a^{\sigma}(t),t)}{\bar{U}^S(t)-\underline{U}^S(t)},$$

where $f(t)$ is type t's prior probability, and $a^{\sigma}(t)$, $a^{\gamma(\sigma)}(t)$ the equilibrium respectively deviating action type t induces. Hence, the ACD of the pooling equilibrium is $\frac{(1-\delta)}{2}\frac{(2-1)}{2-0}+\delta\frac{2-1}{2+\varepsilon-0}=\frac{1}{4}+\delta\frac{2-\varepsilon}{8+4\varepsilon}$. The ACD of the separating equilibrium is $\delta\frac{(2+\varepsilon-2)}{2+\varepsilon-0}=\frac{\delta\varepsilon}{2+\varepsilon}$. The separating equilibrium is always ACDC. In addition, the ACD of the separating equilibrium goes to zero if δ or ε goes to zero.

One may be worried that ACDC always selects the most informative equilibrium. This is not the case, as the following two examples show. First consider

[35] An equilibrium of a game with a one-dimensional type and action set is a partition equilibrium if there exists a partition $t_0 < t_1 < \cdots < t_{n-1} < t_n$ of T such that each type in (t_{i-1}, t_i) induces action a_i with $a_1 < a_2 < \ldots < a_n$. Hence, a partition equilibrium is characterized by a vector $a = (a_1, \ldots, a_n)$ and a partition equilibrium outcome set can be represented by a finite union of compact subsets of $\mathbb{R}^1, \ldots, \mathbb{R}^n$.

Game C in Table 3.3, which is a reproduction of game 2 in Farrell (1993) and example 2 of Matthews, Okuno-Fujiwara & Postlewaite (1991). In this game, the two types occur with equal probability. The game has a separating equilibrium where type t_1 induces a_1 and type t_2 induces a_2. In addition, the game has a pooling equilibrium where both types elicit a_3. Here, the separating equilibrium is not announcement proof because it admits the credible announcement "I am not going to tell you my type." In contrast, no type would want to send an announcement in the pooling equilibrium, which is thus announcement proof. We share the intuition of Farrell and Matthews et al. that the announcement (and neologism) proofness criteria (and hence ACDC) appropriately reject the communication outcome.

TABLE 3.3
GAME C

	a_1	a_2	a_3
$t_1(\frac{1}{2})$	1, 3	0, 0	2, 2
$t_2(\frac{1}{2})$	0, 0	1, 3	2, 2

Game D (Table 3.4) extends the previous game to three Sender types with a twist. For $\varepsilon = 0$, we get an equivalent result as in the previous game. Game D then has a separating equilibrium, where t_i induces a_i and which admits the credible announcement where t_1, t_2 and t_3 deviate to a_4. It also has an announcement proof pooling equilibrium, where all types induce a_4. For small but positive ε, the two equilibria remain intact, but the pooling equilibrium admits a credible announcement where t_1 and t_2 deviate to a_5. Hence, announcement proofness is silent here. ACDC selects the pooling equilibrium: The ACD of the separating equilibrium is $\frac{1}{2} - \frac{1}{6+3\varepsilon}\varepsilon$ and that of the pooling equilibrium $\frac{2}{6+3\varepsilon}\varepsilon$.

TABLE 3.4
GAME D

	a_1	a_2	a_3	a_4	a_5
$t_1(\frac{1}{3})$	1, 3	0, 0	0, 0	2, 2	$2+\varepsilon,\ 2+\varepsilon$
$t_2(\frac{1}{3})$	0, 0	1, 3	0, 0	2, 2	$2+\varepsilon,\ 2+\varepsilon$
$t_3(\frac{1}{3})$	0, 0	0, 0	1, 3	2, 2	$2-\varepsilon,\ 0$

Notes: $0 \le \varepsilon < 1$.

3.4 ACDC versus Other Criteria in a Veto Threats Game

We illustrate the comparative advantage of ACDC over other criteria in a simple continuous external veto threats game (Game E in Table 3.5). At the end of this section, we show that ACDC selects a unique equilibrium in the class of external veto threats games to which this game belongs. The model is close to Matthews' (1989) model, who introduced veto threat games. The difference is that in our model, the disagreement point is external, in the sense that the Sender's payoff does not depend on her type. Whereas in Matthews' model the maximum equilibrium size is two, these games can have a large equilibrium set, with fine partitions and continua of equilibria. This makes them a good testing ground for refinements.

In Game E (Table 3.5), the outcome of the game x is a point on the interval [0,1] or the disagreement point $\delta \notin [0,1]$. The Sender's payoff on the interval depends on t: $U^S(x,t) = -\mid x - t \mid$. The larger the distance between the outcome x and her type t, the lower the Sender's payoff. The Sender's type t is drawn from the uniform distribution on [0,1]. The Receiver's payoff on the interval, $U^R(x) = -x$, is independent of t: he always prefers smaller outcomes to larger ones. It is a veto threats game, because the Sender can veto the Receiver's action, in which case the outcome is the disagreement point δ. The disagreement point payoffs are $U^R(\delta) = -\frac{5}{4}$ and $U^S(\delta,t) = U^S(\delta) = -\frac{1}{5}$. Observe that the Receiver prefers δ to all outcomes on the line larger than $\frac{5}{4}$ and that the Sender prefers δ to all outcomes on the line more than $\frac{1}{5}$ away from her type t.

TABLE 3.5
GAME E

$t \sim$ U[0,1]	$U^R(x) = -x$ for $x \in \mathbb{R}$ $U^S(x,t) = -\mid x - t \mid$ for $x \in \mathbb{R}$ $U^R(\delta) = -\frac{5}{4}$ and $U^S(\delta) = -\frac{1}{5}$

Like Matthews (1989), we look at a refinement of perfect Bayesian equilibrium that restricts the Receiver to pure strategies and lets the Sender take into account that she may tremble at the veto stage (with epsilon probability). As we showed in Chapter 2, in this game all equilibria are partition equilibria which can be characterized by a finite set of Receiver actions $a_1 < a_2 < ... < a_n$, where n is called the size of the equilibrium. Senders simply induce that action closest to their type and accept it if it gives them positive payoff.

The game has a unique pooling equilibrium, in which the Receiver always proposes action $\frac{1}{5}$. This optimal action involves a trade-off between maximizing the probability that the action is accepted and maximizing the payoff of the action once it is accepted. In addition, the game has continua of size-2 and size-3 equilibria. The set of size-2 equilibria is characterized by $\{a_1, a_2 = a_1 + \frac{2}{5}\}$, with $a_1 \in [0, \frac{1}{5}]$, whereas the set of size-3 equilibria is characterized by $\{a_1 = 0, a_2, a_3 = a_2 + \frac{2}{5}\}$ with $a_2 \in (0, \frac{1}{5}]$. In this game, credible neologisms coincide with (weakly and ordinary) credible announcements. Both equilibria admit credible neologisms. All equilibria have a credible neologism 'at the end': $\left\langle \frac{3}{10}, [\frac{2}{5}, 1) \right\rangle$ in the pooling equilibrium and $\left\langle \frac{4}{5}, (\frac{2}{5} + \frac{1}{2} a_n, 1] \right\rangle$ in the other equilibria. If $a_1 > 0$, also a credible neologism 'at the beginning' of the form $\left\langle 0, [0, \frac{1}{2} a_1) \right\rangle$ exists. The game has a unique ACDC equilibrium, which is $\{0, \frac{1}{5}, \frac{3}{5}\}$. The intuition is that this equilibrium does not have a neologism at the beginning and minimizes both the frequency and intensity of deviations at the end by maximizing a_n.

We now turn to other refinements and non-equilibrium solution concepts. From our analysis above, it follows that no equilibrium in Game E is neologism or announcement proof.[36] Rabin & Sobel (1996) propose the recurrent mop criterion, which can select equilibria that, although not impervious to credible deviations, are likely to recur in the long run, because they are frequently deviated to. The authors restrict their definition of the recurrent mop to games with a finite number of actions as it may run into problems in continuous games, amongst others because the deviation correspondence may not converge in these settings. In Game E, it is hard to evaluate the solution concept. In a similar game with just two equilibria (e.g. when $U^R(\delta) = \frac{5}{2}$ and $U^S(\delta) = \frac{1}{4}$) one can show that even if the deviation correspondence would converge, neither equilibrium is stable and both are recurrent.[37]

We now turn to approaches that are not based on credible deviations. The simplest are ex-ante efficiency and influentiality, both of which are not very attractive. Ex-ante efficiency selects the equilibrium with the highest ex-ante payoff for the Sender. One problem with this is that it is not clear how the Sender type can commit ex-post to an ex-ante optimal strategy. Furthermore, which equilibrium is ex-ante optimal is not invariant to affine transformations of payoffs of a subset of types (which should not affect the game). Hence, as long as some types receive less in equilibrium A than in Equilibrium B and the other way around, one can make either equilibrium ex-ante more efficient. In Game E, we can make any size-3 equilibrium efficient by multiplying the payoffs of types around a_2 and a_3 with a suitably large number. (In other games, e.g. with $U^R(\delta) = \frac{5}{2}$ and $U^S(\delta, t) = \frac{1}{4}$, one can also make the pooling equilibrium ex-ante efficient.) Influentiality selects the equilibrium with the largest equilibrium action set. As Games C and D showed, the most influential equilibria need not be the most plausible. Furthermore, Game E has a whole continuum of size-3 equilibria.

The Communication Proofness (CP) criterion of Blume & Sobel (1995) singles out equilibria that would not be destabilized if new opportunities to com-

[36] As in most cheap talk games, neither equilibrium admits strongly credible announcements.
[37] In particular, one can show that for both equilibria it must hold that neither equilibrium lies in the deviation correspondence of that equilibrium.

municate arose. CP looks for partitions that distinguish good and bad equilibria, where in each partition good equilibria cannot be destabilized by other good equilibria. An equilibrium survives CP if it is a good equilibrium in some such partition. In Game E all equilibria are stable according to CP, as in each equilibrium some Sender-type in each partition-element receives her maximum payoff, so that no equilibrium can ever be destabilized.

The No Incentive to Separate (NITS) criterion (Chen, Kartik, & Sobel, 2008) is up till now the only refinement based on some notion of stability that can successfully select an equilibrium in the Crawford-Sobel (1982) setting. NITS starts by specifying a 'lowest type,' a type with the property that all other types prefer to be revealed as themselves rather than as that lowest type. An equilibrium survives NITS if the lowest type has no incentive to separate, i.e. if the lowest type prefers her equilibrium outcome to the outcome she would get if she could reveal her type. In the Crawford-Sobel model, only the maximum size equilibrium outcome satisfies NITS (under some general monotonicity assumption). In Game E, such a 'lowest type' cannot easily be formulated. All types in $[0,\frac{1}{5}]$ are lowest types according to Chen et al.'s definition. Still, one can argue that $t=0$ is a natural lowest type in our game. Under this assumption, the size-2 equilibrium $\{0,\frac{2}{5}\}$ and all size-3 equilibria survive NITS. (By making $U^S(\delta)\,/\,U^R(\delta)$ arbitrarily close to zero, one can make the maximum size arbitrarily high and there will be a NITS equilibrium of each size $2,\ldots,n$.)

Also non-equilibrium concepts exist. Rabin (1990) introduced the concept of Credible Message Rationalizability (CMR). This non-equilibrium concept proposes conditions under which communication can be guaranteed to happen. It assumes that rational players take truth-telling as a focal point, but use the strategic incentives of the game to check whether truth-telling is rational. In Game E, CMR can only guarantee that the 0 type can send a credible message (and is silent about what other types do). CMR requires that all Sender-types who send a credible message receive an action in which they achieve their maximum payoff. This would imply that the Receiver does not best respond to credible messages (of all types except 0), which cannot be the case under CMR.[38]

Blume, Kim & Sobel (1993) put forward the Partial Common Interest (PCI) concept. A partition of the typeset satisfies PCI "if types in each partition element unambiguously prefer to be identified as members of that element, and there is no finer partition with that property." PCI does not make a definite prediction in Game E, as no partition of the type space (except $0 = t_0 < t_1 = 1$) satisfies PCI. The main reason is that the highest Sender-type of a partition-element always prefers the Receiver to believe that the upper boundary is higher than the true boundary (except for types $t = 0$ or $t = 1$). Finally, the 'partition' 0 and $(0, B]$ is not PCI, as 0 (which is the best response if the Sender is 0) is also a best response to some Receiver-beliefs with support on the interval $(0,1]$.

We finish this section by showing that the uniqueness of the ACDC equilibrium is not an artifact of the specific characteristics of Game E. Game E belongs to a wider class of veto threat games introduced in Chapter 2. Here we show that ACDC, under somewhat stricter conditions, selects a unique equilibrium. This result is interesting in its own right, as these games model relevant settings of information transmission under power differences. We assume the Sender's type t is uniformly distributed on the interval [0,1]. We model the player's bargaining power as the payoff of the disagreement point $U^R(\delta)$ and $U^S(\delta)$, where we assume $U^S(\delta, t) = U^S(\delta)$ does not depend on t. U^R and U^S satisfy the following assumptions:

(3.5) U^R on $\mathbb{R}$ is twice continuously differentiable, unimodal with a peak at 0 and concave.

[38] Rabin also introduces an equilibrium version of CMR, Credible Message Equilibria (CME), but as a consequence of the previous analysis, neither equilibrium in Game E can be a CME.

(3.6) $U^S(x,t)$ can be written as a function $f(t-x)$, for all x in $\mathbb{R}$, t in [0,1], where f is continuously differentiable, symmetric, concave, strictly increasing in $\mathbb{R}_-$ and for all $y \in \mathbb{R}$ there is a $z > 0$ such that $f(z) < y$ and $f(-z) < y$; $U^S(\delta) < f(0)$.[39, 40]

In Chapter 2, we show that only partition equilibria exist. Here we show that there is a unique ACDC equilibrium:

Proposition 3.3 *Under assumptions* (3.5) *and* (3.6), *the unique ACDC equilibrium is the maximum size equilibrium with the highest equilibrium action.*

In sum, Game E illustrates that current criteria can fail to be predictive for various reasons. Our intuition is that this is due to the fact that they are based on a dichotomous notion of stability. In contrast, ACDC selects a unique equilibrium in the class of games Game E belongs to.

3.5 Crawford-Sobel Game

In this section, we apply ACDC to select equilibria in the leading uniform-quadratic case of Crawford & Sobel's (1982) cheap talk game (henceforth 'CS game'). In this CS game, types are uniformly distributed on [0,1], the action space is [0,1], $U^R(a,t) = -(a-t)^2$ and $U^S(a,t) = -(a-(t+b))^2$, with $b > 0$ capturing the Sender bias. (The Sender has no veto in this game.)

Crawford and Sobel (1982) show that this game only has (perfect Bayesian) partition equilibria and that the maximum equilibrium size $n(b)$ is the largest integer n for which

[39] Due to the invariance of games (and the ACDC) to affine payoff transformations, it is actually only required that $U^S(x,t)$ can be written as a function $a(t)+b(t)\cdot f(t-x)$, where $b>0$ and f should adhere to the conditions in (3.6) (with $U^S(\delta) < a + b \cdot f(0)$).

[40] Observe that (3.6) implies assumptions (A2)-(A5) in Chapter 2. Our assumptions here are stricter. In particular, they require a uniform type distribution and a symmetric and concave payoff function for the Sender.

$$(3.7) \qquad 2n(n-1)b < 1.$$

In addition, the game has a unique size-n equilibrium for each $n \in \{1,...,n(b)\}$. Let

$$(3.8) \qquad t_i^n \equiv \frac{i}{n} - 2bi(n-i).$$

for $i = 0,...,n$ and $n = 1,...,n(b)$. In the size-n equilibrium, types in $[t_{i-1}^n, t_i^n)$ send the same equilibrium message, which induces the Receiver to choose action

$$(3.9) \qquad a_i^n = \frac{1}{2}(t_{i-1}^n + t_i^n),\ i = 1,...,n.$$

We start by deriving all credible neologisms the equilibria admit. For each credible neologism $\langle \tilde{a}, N \rangle$, the set of deviating types N turns out to be an interval between some $\underline{\tau}$ and $\overline{\tau}$. Hence, we can characterize neologisms by $[\underline{\tau}, \overline{\tau}]$ alone, since the Receiver's best response is $\tilde{a} = \dfrac{\underline{\tau} + \overline{\tau}}{2}$. An equilibrium can admit three types of credible neologisms. First of all, there may be a credible neologism which includes $t = 0$. If this credible neologism exists, then it has the shape $[0, \overline{\tau}_0^n)$ where

$$\overline{\tau}_0^n = \frac{2}{3}a_1^n - \frac{4}{3}b = \frac{1}{3n} - \frac{2}{3}b(n+1).$$

Chen, Kartik & Sobel (2008) show that an equilibrium that fails NITS has a credible neologism of this kind. They also prove that only the size-$n(b)$ equilibrium satisfies NITS, so that the credible neologism $[0, \overline{\tau}_0^n)$ exists if and only if $n < n(b)$.

Second, Farrell (1993) shows that if $b < \frac{1}{2}$, the game has a credible neologism on the right-end of the type space of the form $(\underline{\tau}_n^n, 1]$ where

$$\underline{\tau}_n^n = 1 - \frac{1}{3n} - \frac{2}{3}b(n+1).$$

Finally, if $n \in \{2, ..., n(b)-1\}$, there are $n-1$ credible neologisms "in the middle." These take the form $(\underline{\tau}_i^n, \overline{\tau}_i^n)$ for $i = 1, ..., n-1$, where $\underline{\tau}_i^n$ $[\overline{\tau}_i^n]$ is indifferent between the equilibrium action a_i^n $[a_i^{n+1}]$ and the neologism action $\tilde{a}_i^n = (\underline{\tau}_i^n + \overline{\tau}_i^n)/2$. We obtain:

$$\begin{aligned} \underline{\tau}_i^n &= \frac{3}{4}a_i^n + \frac{1}{4}a_{i+1}^n - 2b \text{ and} \\ \overline{\tau}_i^n &= \frac{1}{4}a_i^n + \frac{3}{4}a_{i+1}^n - 2b, \end{aligned} \tag{3.10}$$

$i = 1, ..., n-1$. If $n = n(b)$, the game has the same types of credible neologisms "in the middle," with the exception that the neologism $\left(\underline{\tau}_1^{n(b)}, \overline{\tau}_1^{n(b)}\right)$ need not exist.[41] Observe that $\overline{\tau}_{i-1}^n < \underline{\tau}_i^n$ for $i = 1, ..., n$, so that none of the credible neologisms overlap. Figure 3.1 illustrates the results for $b = \frac{1}{18}$.

It seems intuitive that the highest size equilibrium is the ACDC equilibrium, since the deviations seem to get smaller and smaller as the size increases. This indeed turns out to be the case. Although one can obtain analytical results for the ACD for specific parameter values, finding the ACDC equilibrium for general b defies an analytical approach. Hence, we calculated the ACD for a very fine grid of b and obtain the following result.

[41] If (and only if) $2bn(b)^2 \geq 1$, there is no credible neologism of the form $(\underline{\tau}_1^{n(b)}, \overline{\tau}_1^{n(b)})$ because $\underline{\tau}_1^{n(b)} = \frac{3}{4}a_1^{n(b)} + \frac{1}{4}a_2^{n(b)} - 2b = -\frac{3}{4n(b)}\left(2bn(b)^2 - 1\right) \leq 0$, which is inconsistent with all types being in the interval [0,1] or the interval $\left(\underline{\tau}_1^{n(b)}, \overline{\tau}_1^{n(b)}\right) = \left(0, t_1^{n(b)}\right)$ being a neologism.

Proposition 3.4 *For all* $b \in \left\{\frac{1}{10000}, \frac{2}{10000}, \ldots, \frac{1}{4}\right\}$ *it holds that the ACD of the size-*n *equilibrium in the CS game is decreasing in* n.

Corollary 3.1 *For all* $b \in \left\{\frac{1}{10000}, \frac{2}{10000}, \ldots, \frac{1}{4}\right\}$, *the size-*$n(b)$ *equilibrium is the unique ACDC equilibrium.*

We also derive the following property of the maximum size equilibrium (for which we do not need to calculate the ACD's for each b):

Proposition 3.5 *The ACD of the size-*$n(b)$ *equilibrium tends to zero if* b *tends to zero in the CS game.*

Hence, the ACD of this equilibrium converges to zero if b approaches zero, i.e. if the interests of the players are almost perfectly aligned. This finding is intuitive because the Sender obtains almost her ideal outcome when b is close to zero, so she will not gain much in the case of deviation, and even if she deviates, the deviation will hardly change the equilibrium.

Proposition 3.4 is in line with NITS (Chen, Kartik & Sobel, 2008), which also selects the size-$n(b)$ equilibrium. If the bias parameter is large, however, the maximum size equilibrium can still have a large ACD, so it may not be all that stable. The prediction of ACDC that the maximum size equilibrium becomes more stable as b becomes smaller (Proposition 3.5) is not made by previous concepts. NITS does not predict this, as it assumes that no type separates if the lowest type does not separate.

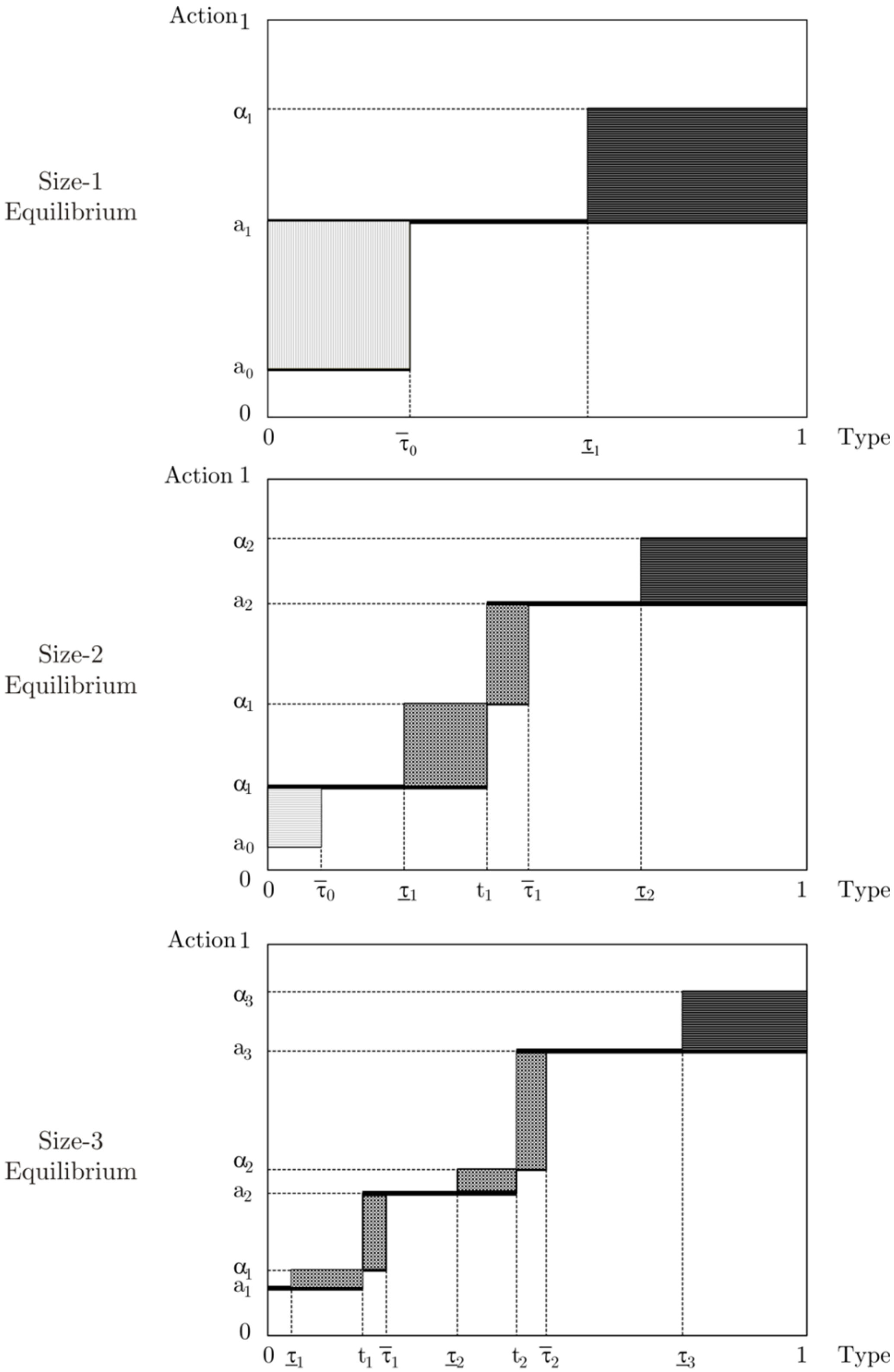

Figure 3.1

The size-1, size-2 and (maximum) size-3 equilibria with the credible neologisms they admit for $b=\frac{1}{18}$. The area of the neologisms give an impression of their contribution to the ACD, although there height contributes quadratically to the ACD.

3.6 Experimental Evidence

In this section, we look at the experimental support for ACDC.

3.6.1 Discrete games

Blume, DeJong, Kim & Sprinkle (2001) provide an experimental analysis of 4 discrete cheap talk games, in which they compare the predictive power of refinements as neologism proofness, influentiality and ex-ante efficiency with PCI. (Credible announcements coincide with credible neologisms in Blume et al.'s games.) They find that PCI is a reliable predictor of when communication takes place and that the equilibrium refinements sometimes but not always improve on PCI. In their Games 1 and 3, the predictions of PCI and announcement proofness (and ACDC) are very much aligned, and borne out by the data. In their Game 2 (see Table 3.6) announcement proofness predicts complete separation while the finest partition consistent with PCI entails partial separation. The data are in line with separation, as a clear majority of 88% of the outcomes is consistent with the separating equilibrium. One could argue that this result does not contradict PCI, because PCI allows multiple patterns including separation (see their footnote 10). As the authors note (in footnote 19), one needs to add neologism proofness to PCI to actually predict that separation happens.

TABLE 3.6
REPRODUCTION OF GAMES 2 AND 4 OF BLUME ET AL. (2001)

	a_1	a_2	a_3	a_4	a_5
t_1	800, 800	100, 100	0, 0	500, 500	0, 400
t_2	x, 100	y, 800	0, 0	500, 500	0, 400
t_3	0, 0	0, 0	500, 800	0, 0	0, 400

Notes: All the three types $\{t_1, t_2, t_3\}$ of the Sender are equally likely and the Receiver can implement one of the actions $\{a_1, ..., a_5\}$. Entry i,j, represents $U^S(t_i, a_j), U^R(t_i, a_j)$. Games 2 and 4 are identical, except that $x = 100, y = 300$ in game 2, whereas $x = 300, y = 100$ in game 4.

Their Game 4 (Table 3.6) is interesting because no equilibrium is announcement proof while PCI makes a prediction. This game has two equilibrium outcomes. Besides the pooling equilibrium where action a_5 is induced there is a partially separating equilibrium where types t_1 and t_2 send a common message that differs from the message of t_3. Types t_1 and t_2 induce a_4 while type t_3 induces a_3. Full separation is not an equilibrium because t_2 prefers to mimic t_1. None of the equilibria satisfies announcement proofness. PCI predicts meaningful communication because the finest partition consistent with PCI is given by $\{\{t_1, t_2\}, \{t_3\}\}$. The partially separating equilibrium only has a credible announcement where t_1 deviates to a_1. Thus, its ACD equals $\frac{1}{3}\frac{(800-500)}{800} = \frac{1}{8}$. The pooling equilibrium admits the credible announcement where t_1 and t_2 deviate to a_4 and t_3 deviates to a_3. Consequently, its ACD is $\frac{1}{3}\left(\frac{(500-0)}{800} + \frac{(500-0)}{500} + \frac{(500-0)}{500}\right) = \frac{7}{8}$. So ACDC predicts that the partially separating equilibrium will be the most observed equilibrium outcome but that it will not be completely stable.

In line with this prediction, Blume et al. find that 37% of the outcomes are consistent with the partially separating equilibrium but no outcome is consistent with the pooling equilibrium. Thus, of the two equilibria, the one with the lowest ACD performs best. Consistent with the ACD measures, much fewer outcomes are in line with the equilibrium selected by ACDC in game 4 than in game 2. In line with the fact that types t_1 have a credible announcement, they turn out to be the ones that are able to credibly identify themselves.

Our conclusion is that our ACDC concept improves the predictions of announcement proofness and that it does at least as well as PCI in explaining the data of Blume, DeJong, Kim & Sprinkle (2001). The extra mileage for ACDC comes from continuous games like the Crawford-Sobel game and the veto threat game. PCI fails to predict any communication at all in these settings, while in accordance with ACDC subjects are able to communicate meaningfully to a large extent, as we will see next.

3.6.2 Crawford-Sobel game

Several experimental studies have been done in discrete versions of the CS game. Cai & Wang (2006) test a discrete version of this game where the Sender's type is uniformly drawn from {1, 3, 5, 7, 9}. They report the results of four treatments that differ in the disalignment between the Sender's and Receiver's preferences. The treatment with the smallest disalignment has the full range of equilibria from pooling to fully separating, while the treatment with the largest disalignment parameter only allows for the pooling equilibrium. In the two treatments in between, the most informative equilibrium is a size- 2 equilibrium. Both NITS and ACDC select a most informative equilibrium in each treatment.[42]

The experimental results are relatively closest to the most informative equilibrium. Except in the case where complete separation is supported in equilibrium, subjects over-communicate compared to the most informative equilibrium, and overcommunication increases as the bias parameter increases. Hence, in agreement with ACDC, behavior departs more from the most informative equilibrium, as its ACD increases.[43]

3.6.3 Veto Threats Game

In Chapter 4, we present new experimental data. We test the predictions of ACDC in five games belonging to the class of veto threat games discussed in section 3.4. We find that ACDC organizes the data well. The ACDC equilibrium performs best, even if it admits credible deviations. Furthermore, in comparable games, the ACDC equilibrium performs better as its ACD decreases. Finally, in

[42] In their treatment with the disalignment parameter equal to 1.2 there is an additional most informative equilibrium $\{1\}$, $\{3579\}$, besides the reported most informative equilibrium $\{13\},\{579\}$. Both equilibria are NITS, while ACDC selects the latter equilibrium. The data are not sufficiently informative to discriminate between these equilibria.

[43] The results of Dickhaut, McCabe & Mukherji (1995) on a Crawford-Sobel game are similar to those reported by Cai and Wang, although they do not interpret their results in terms of overcommunication. More recently, Wang, Spezio & Camerer (2010) replicate the results of Cai & Wang (2006) and find that look-up patterns of Senders (as measured by eye-tracking) reveals a significant amount of information about their type.

a setting with continua of equilibria similar to Game E, we find that the closer equilibria are to the ACDC equilibrium, the better they perform.

3.7 Conclusion

This chapter generalizes refinements based on credible deviations, in particular neologism proofness and announcement proofness. We started with an intuition for why the frequency and size of credible deviations could affect equilibrium stability in a continuous rather than a binary manner. Consequently, we formalized this intuition in ACDC, which measures the (in)stability of cheap talk equilibria and determines which are most plausible. We show an ACDC equilibrium exists under general conditions unlike existing concepts. Furthermore, the predictions of ACDC are meaningful in previously analyzed settings and organize the data of previous experiments well. Finally, ACDC makes predictions in settings where other concepts cannot.

In Chapter 4, we find support for ACDC in a new experimental setting where other criteria remain silent.

3.8 Appendix: Proofs

3.8.1 Proposition 3.2

Proof of Proposition 3.2 $ACD_\gamma(s)$ achieves a minimum on each compact subset S_i and hence achieves a minimum on S. As a consequence, also $\min_{\Sigma^*} ACD_\gamma(\sigma)$ is nonempty, so that at an ACDC equilibrium exists. *Q.E.D.*

3.8.2 Proposition 3.3

For the proof of Proposition 3.3, we introduce some definitions and results from Chapter 2, and derive two helpful lemmas.

Observe that in this game, a neologism $\langle \tilde{a}, N \rangle$ is credible relative to equilibrium σ^* if and only if

$$\tilde{a} \in \arg\max_{a\in\mathbb{R}} P\left\{U^S(a,t) \geq 0 \middle| t \in N\right\}\left(U^R(a) - U^R(\delta)\right), \text{ and}$$

$$\text{for all } k = 1,\ldots,n \text{ it holds that } t \in [t_{k-1}, t_k] \cap N \Rightarrow U^S(\tilde{a},t) > U^S(a_k,t) \text{ and}$$

$$t \in [t_{k-1}, t_k] \setminus N \Rightarrow U^S(\tilde{a},t) \leq U^S(a_k,t).$$

Lemma 3.1 If $\langle \tilde{a}, N \rangle$ *is a credible neologism relative to equilibrium* σ^*, *then* N *is an interval.*

Proof. The proof is by contradiction. Suppose $0 \leq t^1 < t^2 < t^3 \leq 1$, $t^1, t^3 \in N$ and $t^2 \notin N$. Suppose further that in equilibrium, type t^i obtains action a^i, $i = 1,2,3$. The fact that the a type's utility is strictly decreasing in the distance between $t - a$ implies $a^1 \leq a^2 \leq a^3$. If $\tilde{a} \leq t^2$ then it must be the case that $\tilde{a} \leq a^2$ (otherwise type t^2 would prefer $\tilde{a}$ over a^2). As a consequence, $\tilde{a} \leq t^3 \leq a^3 = a^2$ because type t^3 must prefer $\tilde{a}$ over a^3 and a^3 over a^2. A contradiction is established, because the fact that the indifference points $t - d$ and $t + d$ are strictly increasing in t implies that type t^2 strictly prefers $\tilde{a}$ over a^2. This is in conflict with the definition of a credible neologism. Analogously, $\tilde{a} > t^2$ can be ruled out, so that N is an interval. *Q.E.D.*

From (3.6), it follows that there is a $d > 0$ such that for all t and $a \in \mathbb{R}$, $U^S(a,t) \geq U^S(\delta)$ if and only if $a \in [t-d, t+d]$. Hence, $t-d$ and $t+d$ are the Sender's indifference points as to whether she accepts action a. From Lemma's 1 and 2 in Lemma 2.1 and Lemma 2.2 it follows that in equilibrium

$$a_1 \geq 0, \quad t_{k-1} - d < a_k \leq t_k - d \quad \textit{for all } k = 2,\ldots,n \textit{ and } t_{k-1} + d \leq a_k \quad \textit{for } k = 3,\ldots,n. \tag{3.11}$$

We can now show that under (3.6), it holds that

Lemma 3.2 *In equilibrium,* $a_k + d = t_k = a_{k+1} - d$ *for* $k = 2,...,n-1$.

Proof. Due to the t being uniformly distributed and (3.6), the indifference points $t - d$ and $t + d$ are uniformly distributed as well. This means that if the Receiver receives a message that identifies Sender types to be in the interval $[t_k, t_{k+1}]$ ($k = 0,...,n-1$), the probability the Sender accepts an action is not higher for an action $a > t_k + d$ than for action $a' = t_k + d$, while $U^R(a) < U^R(a')$. Hence, for the equilibrium action a_k it holds true that $a_k \le t_{k-1} + d$ and by (3.11), this means $a_k = t_{k-1} + d \le t_k - d$ for $k = 3,...,n$. Now, suppose that $t_{k-1} + d < t_k - d$ for some $k = 3,...,n$. This means that $a_k < t_k - d$ and hence $U^S(a_k, t_k) < 0$. Since $U^S(a_k, t_k) = U^S(a_{k+1}, t_k)$, this implies, however, that $a_{k+1} > t_k + d$, which for $k = 3,...,n-1$ is a contradiction with $a_k \le t_{k-1} + d$ for $k = 3,...,n$. Hence, $a_k = t_{k-1} + d = t_k - d$ for $k = 3,...,n-1$. Consequently, $a_k + d = t_k = a_{k+1} - d$ for $k = 3,...,n-1$.

Furthermore, from the discussion above we have that $t_2 = a_3 - d$ and that $a_2 \le t_1 + d$. In addition, from (3.11) it follows that $a_2 \le t_2 - d$. Hence, a necessary condition on a_2 is that $a_2 \in \arg\max_{t_1 + d \le a \le t_2 - d} \left(U^R(a) - U^R(\delta)\right)\left(a + d - t_1\right)$. Analogously to our discussion in the proof of Proposition 2.2, one can show that this implies that a_2 must be equal to $t_2 - d$. As a result, $a_2 + d = t_2 = a_3 - d$. *Q.E.D.*

Proof of Proposition 3.3 Suppose that the game has more than one equilibrium outcome. If $\overline{x} \le 2d$, then consider the equilibrium outcome σ^* with $a_1 = 0$ and a_2 such that $a_2 \in \arg\max_{a \in \mathbb{R}} U^R(a)\left(\min\{a + d, 1\} - \frac{1}{2}a_2\right)$. If $\overline{x} > 2d$, let n be the natural number for which $\overline{x} - 2dn \le 0$ and $\overline{x} - 2d(n-1) > 0$, and consider the following $\sigma^* : a_1 = 0$; $a_k = \overline{x} - 2d(n-k-2)$, $k = 2,...,n$. We now show that σ^* has the maximum equilibrium size and is the unique ACDC equilibrium outcome.

From (2.2) and (3.6), it follows that there exists an $\overline{x} \in \mathbb{R}$ such that

$$\begin{aligned} &U^R(x) - U^R(\delta) + 2dU^{R\prime}(x) \geq 0 \text{ for all } x \in [0, \overline{x}) \text{ and} \\ &U^R(x) - U^R(\delta) + 2dU^{R\prime}(x) < 0 \text{ for all } x \in (\overline{x}, 1-d]. \end{aligned} \tag{3.12}$$

where a prime ($'$) denotes a derivative with respect to x. Let a^* denote the highest equilibrium action a_n in σ^*. Using (3.12), it can be verified that σ^* constitutes the highest size equilibrium, analogously to the proof of Proposition 2.2. Similarly, it can be verified that the highest action a^{**} in any other equilibrium σ^{**} must be smaller than a^*:

$$a^{**} \leq a^* \leq 1 - d.$$

If $a^{**} < 1-d$, σ^{**} has at least one credible neologism: Types in the interval $(\underline{\tau}^{**}, 1]$ are willing to send a credible neologism $\left\langle \tilde{a}^{**}, (\underline{\tau}^{**}, 1] \right\rangle$, where

$$\underline{\tau}^{**} = \frac{1}{2}\left(a^{**} + \tilde{a}^{**}\right), \text{ and}$$

$$\tilde{a}^{**} \in \underset{a \in (a^{**}, \lambda(1)]}{\arg\max} \left(U^R(a) - U^R(\delta)\right) \frac{a + d - \underline{\tau}^{**}}{1 - \underline{\tau}^{**}}.$$

To proof that σ^* is an ACDC equilibrium, we first show it has at most one credible neologism (claim 1) and this credible neologism, if it exists, maximizes $\underline{\tau}^{**}$ and minimizes $\tilde{a}^{**} - a^{**}$ (claim 2).

In order to prove claim 1, suppose that σ^* has another credible neologism. By Lemma 3.1, the set of types that send the credible neologism relative to equilibrium σ^* is an interval. We can exclude neologisms that induce the Receiver to propose $a = 0$, because $a_1 = 0$ is already an equilibrium action. Hence, the neologism $\tilde{a}$ (with supremum neologism type $\tilde{\tau}$) is in between two equilibrium actions a_{k-1} and a_k. Due to Lemma 3.1, $a_{k-1} < \tilde{a} < \tilde{\tau} < a_k$. This implies that $U^S(\tilde{a}, \tilde{\tau}) \leq 0$, because if $U^S(\tilde{a}, \tilde{\tau}) > 0$, action $\tilde{\tau} - d$ would be better for the

Receiver than $\tilde{a}$ after receiving the neologism. Consequently, $U^S(a_{k-1},\tilde{\tau}) < U^S(\tilde{a},\tilde{\tau}) \le 0$ and $U^S(a_k,\tilde{\tau}) < U^S(\tilde{a},\tilde{\tau}) \le 0$. This means that an $\varepsilon > 0$ exists such that a types in $(\tilde{\tau}-\varepsilon,\ \tilde{\tau}+\varepsilon)$ receive 0 payoff in equilibrium. Since this is not the case in σ^*, σ^* has no other neologisms.

The proof of claim 2 proceeds as follows. Note that $\tilde{a}^{**} = \min\{\overline{a}^{**}, 1-d\}$, where $\overline{a}^{**} = \arg\max_{a\in\mathbb{R}} \left(U^R(a) - U^R(\delta)\right)\dfrac{a+d-\underline{\tau}^{**}}{1-\underline{\tau}^{**}}$. We know $\overline{a}^{**} > a^{**}$, because the solution to $\arg\max_{a\in\mathbb{R}} \left(U^R(a) - U^R(\delta)\right)\dfrac{a+d-\underline{t}}{1-\underline{t}}$ is increasing in $\underline{t}$ and a^{**} is the solution for $\underline{t} = t_{n-1}$, and $\overline{a}^{**}$ is the solution to the problem with $\underline{t} \ge a^{**} > t_{n-1}$. Moreover,

$$U^R(\overline{a}^{**}) - U^R(\delta) + U^{R\prime}(\overline{a}^{**})\left(\overline{a}^{**} + d - \underline{\tau}^{**}\right) = U^R(\overline{a}^{**}) - U^R(\delta) + U^{R\prime}(\overline{a}^{**})\left(\frac{\overline{a}^{**} - a^{**}}{2} + d\right)$$

$= 0$ implies that

$$\overline{a}^{**} - a^{**} = -2\frac{U^R(\overline{a}^{**})}{U^{R\prime}(\overline{a}^{**})} - 2d.$$

From the concavity of U^R it follows that $\dfrac{U^R(a)}{U^{R\prime}(a)}$ is increasing in a. Hence, $\overline{a}^{**} - a^{**}$ is decreasing in a^{**}. In particular, this implies that $\tilde{a}^{**} - a^{**}$ is decreasing in a^{**}. Moreover, $\underline{\tau}^{**}$ is increasing in a^{**}.

Finally, to show that σ^* is an ACDC equilibrium, we show that it has the lowest ACD. By Lemma 3.2, for equilibrium σ^{**} it must then hold that $a_1^{**} > 0$ or $a^{**} < a^*$. If $a_1^{**} > 0$, then a neologism $\langle \tilde{a}_0, [0,\overline{\tau}_0]\rangle$ exists with $\tilde{a}_0 < a_1^{**}$.[44] Suppose now that $a^{**} < a^*$. If σ^* does not admit a credible neologism, it is

[44] If $a_1^{**} \ge 2d$, $\tilde{a}_0 = d$ and $U^S(d,\tilde{\tau}_0) = U^S(a_1^{**},\tilde{\tau}_0)$. If $a_1^{**} \le d$, $\tilde{a}_0 = 0$ and $U^S(\tilde{a}_0,\tilde{\tau}_0) = U^S(0,\tilde{\tau}_0)$. If $d < a_1^{**} < 2d$, $\tilde{a}_0 = \tilde{\tau}_0 + d$ and $U^S(\tilde{a}_0,\tilde{\tau}_0) = U^S(a_1^{**},\tilde{\tau}_0)$. This has a solution, because $U^S(\tilde{\tau}_0 - d,\tilde{\tau}_0) - U^S(a_1^{**},\tilde{\tau}_0) > 0$ for $\tilde{\tau}_0 = 0$ and $U^S(\tilde{\tau}_0 - d,\tilde{\tau}_0) - U^S(a_1^{**},\tilde{\tau}_0) < 0$ for $\tilde{\tau}_0 = a_1^{**}$.

evident that $ACD(\sigma^*) = 0 < ACD(\sigma^{**})$. Hence, suppose that σ^* admits the credible neologism $\langle \tilde{a}^*, [\underline{\tau}^*, 1] \rangle$.

We can now compare the ACD of σ^* and σ^{**}. First, $CD^{\sigma^*}(t) = 0$ for $t \in [0, \underline{\tau}^*)$. Second, we show that $U^S(\tilde{a}^{**}, t) - U^S(a^{**}, t) > U^S(\tilde{a}^*, t) - U^S(a^*, t)$ for $t \in [\underline{\tau}^*, a^{**} + d)$. Due to claim 2 $\tilde{a}^{**} - a^{**} > \tilde{a}^* - a^*$ and $\underline{\tau}^{**} < \underline{\tau}^*$. If $t \leq \tilde{a}^{**} < \tilde{a}^*$, then $U^S(a^{**}, t) < U^S(a^*, t)$ and $U^S(\tilde{a}^{**}, t) > U^S(\tilde{a}^*, t)$, so that the result is immediate. Assume now that $\tilde{a}^{**} < t$. By (3.6), $U^S(a, t)$ is concave in a, such that for $x < y \leq t$ and $b, c > 0$ it holds that:

$$U^S(y,t) - U^S(x,t) \leq U^S(y-b,t) - U^S(x-b,t) < U^S(y-b,t) - U^S(x-b-c,t).$$

Hence, for $t \in [\underline{\tau}^*, \tilde{a}^*]$ we have that $U^S(\tilde{a}^*, t) - U^S(a^*, t) \leq U^S(t,t) - U^S(a^*, t) \leq U^S(\tilde{a}^{**}, t) - U^S(a^* - t + \tilde{a}^{**}, t) < U^S(\tilde{a}^{**}, t) - U^S(a^{**}, t)$. (Observe that $t - a^* < \tilde{a}^* - a^* < \tilde{a}^{**} - a^{**}$.) Similarly, for $t \in (\tilde{a}^*, a^{**} + d]$, $U^S(\tilde{a}^*, t) - U^S(a^*, t) \leq U^S(\tilde{a}^{**}, t) - U^S(a^* - \tilde{a}^* + \tilde{a}^{**}, t) < U^S(\tilde{a}^{**}, t) - U^S(a^{**}, t)$. As a consequence, $CD^{\sigma^{**}}(t) > CD^{\sigma^*}(t)$ for $t \in [\underline{\tau}^{**}, a^{**} + d)$. Finally, $CD^{\sigma^{**}}(t) = 1 \geq CD^{\sigma^*}(t)$ for $t \in [a^{**} + d, 1]$. Together, this implies that $ACD(\sigma^{**}) = E_t\left[CD^{\sigma^{**}}(t)\right] > E_t\left[CD^{\sigma^*}(t)\right] = ACD(\sigma^*)$.

In sum, if σ^{**} is different from σ^*, then either $a_0^{**} > 0$ or $a^{**} < a^*$ and in both cases $ACD(\sigma^{**}) > ACD(\sigma^*)$. Therefore, σ^* is the unique ACDC equilibrium. *Q.E.D.*

3.8.3 Proposition 3.4 and Proposition 3.5

We did the analysis in the following way. First, we obtain closed-form solutions for the ACD for each value of b and second we calculate the ACD for a fine grid of b.

The ACD of equilibrium σ in the CS game is equal to

$$ACD(\sigma) = E_t\left[\frac{U^S(\tilde{a}^\sigma(t),t) - U^S(a^\sigma(t),t)}{\bar{U}^S(t) - \underline{U}^S(t)}\right]$$

$$= \int_0^1 \frac{U^S(\tilde{a}^\sigma(t),t) - U^S(a^\sigma(t),t)}{U^S(\min\{t+b,1\},t) - \min\{U^S(0,t),U^S(1,t)\}} dt$$

$$= \int_0^1 \frac{\left(a^\sigma - (t+b)\right)^2 - \left(\tilde{a}^\sigma - (t+b)\right)^2}{-\max\{0,t+b-1\}^2 + \max\{(t+b)^2,(t+b-1)^2\}} dt.$$

Note that $(t+b-1)^2 > (t+b)^2$ if and only if $t < \frac{1}{2} - b$. Suppose $a^\sigma(t)$ and $\tilde{a}^\sigma(t)$ are constant and $\bar{U}^S(t) = 0$ on the interval $[\underline{t},\overline{t}]$. Let $\hat{t} \equiv \max\{\underline{t},\min\{\overline{t},\frac{1}{2} - b\}\}$. Then, $\int_{\underline{t}}^{\overline{t}} CD(t,\sigma)dt$ is equal to

$$h(b,a^\sigma,\alpha^\sigma,\underline{t},\overline{t}) \equiv \int_{\underline{t}}^{\overline{t}} \frac{\left(a^\sigma - (t+b)\right)^2 - \left(\tilde{a}^\sigma - (t+b)\right)^2}{\max\{(t+b)^2,(t+b-1)^2\}} dt$$

$$= \int_{\underline{t}}^{\hat{t}} \frac{\left(a^\sigma - (t+b)\right)^2 - \left(\tilde{a}^\sigma - (t+b)\right)^2}{(t+b-1)^2} dt + \int_{\hat{t}}^{\overline{t}} \frac{\left(a^\sigma - (t+b)\right)^2 - \left(\tilde{a}^\sigma - (t+b)\right)^2}{(t+b)^2} dt$$

$$= (a^\sigma - \tilde{a}^\sigma)\left(\frac{(a^\sigma + \tilde{a}^\sigma - 2)(\hat{t} - \underline{t})}{(b-1+\hat{t})(b-1+\underline{t})} + 2\log\left[\frac{b-1+\underline{t}}{b-1+\hat{t}}\right]\right)$$

$$+(a^\sigma - \tilde{a}^\sigma)\left(\frac{(a^\sigma + \tilde{a}^\sigma)(\overline{t} - \hat{t})}{(b+\overline{t})(b+\hat{t})} + 2\log\left[\frac{b+\hat{t}}{b+\overline{t}}\right]\right).$$

As noted before, an equilibrium of size n can have a neologism in the beginning $\tilde{a}_0^n$, a neologism at the end $\tilde{a}_n^n$ and at most $n-1$ neologisms in the middle, $\tilde{a}_i^n, i = 1,...,n-1$. The size-1 equilibrium has a neologism at the beginning and at the end. The maximum size $n(b)$ equilibrium has a neologism at the end and neologisms in the middle $\tilde{a}_i^n, i = \underline{i}(b),...,n-1$, where $\underline{i}(b) = 1$ if $2bn(b)^2 < 1$ and $\underline{i}(b) = 2$ if $2bn(b)^2 \geq 1$. Size-n equilibria with $1 < n < n(b)$ admit all neologisms specified above. Observe that $\tau_{n-1}^n < 1-b$, such that $\bar{U}^S(t) = 0$ except for the highest types of the highest neologism, such that $h(b,a^\sigma,\tilde{a}^\sigma,\underline{t},\overline{t})$ can be used to calculate the contribution to the ACD for neologisms $\underline{i}(b) = 1,...,n-1$. For the highest neologism, the contribution to the ACD is equal to

$$\begin{aligned}\overline{h}(b,n) &= h(b,a_n^n,\tilde{a}_n^n,\underline{\tau}_n^n,b-1)+\int_{b-1}^{1}\frac{\left(a_n^n-(t+b)\right)^2-\left(\tilde{a}_n^n-(t+b)\right)^2}{-(t+b-1)^2+(t+b)^2}dt\\ &= h(b,a_n^n,\tilde{a}_n^n,\underline{\tau}_n^n,b-1)+\frac{1}{2}(a_n^n-\tilde{a}_n^n)(a_n^n+\tilde{a}_n^n-1)\log[2b+1].\end{aligned}$$

Let σ_b^n be the size-n equilibrium of the game with bias parameter b. Then, the ACD of the pooling equilibrium is

$$ACD(\sigma_b^1)=h(b,a_1^1,\tilde{a}_0^1,0,\overline{\tau}_0^1)+\overline{h}(b,1).$$

The ACD of the maximum-size equilibrium is

$$ACD\left(\sigma_b^{n(b)}\right)=\sum_{i=\underline{i}(b)}^{i=n(b)-1}[h(b,a_i^{n(b)},\tilde{a}_i^{n(b)},\underline{\tau}_i^{n(b)},t_i^{n(b)})+h(b,a_{i+1}^{n(b)},\tilde{a}_i^{n(b)},t_i^{n(b)},\overline{\tau}_i^{n(b)})]+\overline{h}(b,n(b)).$$

The ACD of a size-n equilibrium with $1<n<n(b)$ is equal to

$$ACD(\sigma_b^n)=h(b,a_1^n,\tilde{a}_0^n,0,\overline{\tau}_0^n)+\sum_{i=1}^{i=n-1}[h(b,a_i^n,\tilde{a}_i^n,\underline{\tau}_i^n,t_i^n)+h(b,a_{i+1}^n,\tilde{a}_i^n,t_i^n,\overline{\tau}_i^n)]+\overline{h}(b,n)$$

Proof of Proposition 3.4 For each $b\in\left\{\frac{1}{10000},\frac{2}{10000},\ldots,\frac{1}{4}\right\}$, one can calculate the (closed-form) value of $ACD(\sigma_b^n)$ for all $1\le n\le n(b)$, and verify that the ACD of the size-n equilibrium in the CS game is decreasing in n.

Proof of Proposition 3.5 Let $\sigma(b)\equiv\sigma_b^{n(b)}$ be the maximum size equilibrium for b. Then,

$$\lim_{b\downarrow 0}ACD(\sigma(b))\le\lim_{b\downarrow 0}E_t\left[\frac{U^S(\tilde{a}^{\sigma(b)}(t),t)-U^S(a^{\sigma(b)}(t),t)}{\min_{t\in T}\{\overline{U}^S(t)-\underline{U}^S(t)\}}\right]\le\lim_{b\downarrow 0}E_t\left[\frac{0-U^S(a^{\sigma(b)}(t),t)}{\frac{1}{4}}\right]$$

$$= -4 \cdot \lim_{b \downarrow 0} EU^S \overset{1}{=} 4 \cdot \lim_{b \downarrow 0} \left(b^2 + \frac{1}{12 n(b)^2} + \frac{b^2 (n(b)^2 - 1)}{3} \right)$$

$$\leq 4 \cdot \lim_{b \downarrow 0} \left(b^2 + \frac{1}{n(b)^2} + b^2 n(b)^2 \right) \overset{2}{\leq} 4 \cdot \lim_{b \downarrow 0} \left(b^2 + \frac{4}{\left(\sqrt{2/b+1} - 1\right)^2} + \frac{\left(b + \sqrt{2b + b^2}\right)^2}{4} \right) = 0$$

Equality 1 follows from the specification of EU^S in Crawford & Sobel (1982). Inequality 2 follows from $n(b) = \left\lceil \frac{1}{2} + \frac{1}{2}\sqrt{2/b+1} \right\rceil - 1$ due to (3.7). The other manipulations are straightforward. *Q.E.D.*