

de Groot Ruiz, Adrian; Offerman, Theo; Onderstal, Sander

**Working Paper**

## Equilibrium Selection in Cheap Talk Games: ACDC rocks when Other Criteria remain silent

Tinbergen Institute Discussion Paper, No. 11-037/1

**Provided in Cooperation with:**

Tinbergen Institute, Amsterdam and Rotterdam

*Suggested Citation:* de Groot Ruiz, Adrian; Offerman, Theo; Onderstal, Sander (2011) : Equilibrium Selection in Cheap Talk Games: ACDC rocks when Other Criteria remain silent, Tinbergen Institute Discussion Paper, No. 11-037/1, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:

<https://hdl.handle.net/10419/86952>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

TI 2011-037/1  
Tinbergen Institute Discussion Paper



# Equilibrium Selection in Cheap Talk Games: ACDC Rocks when Other Criteria Remain Silent

*Adrian de Groot Ruiz*

*Theo Offerman\**

*Sander Onderstal\**

*University of Amsterdam.*

*\* Tinbergen Institute.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900  
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 8579

# Equilibrium Selection in Cheap Talk Games: ACDC Rocks When Other Criteria Remain Silent\*

Adrian de Groot Ruiz<sup>a</sup>, Theo Offerman<sup>b</sup> and Sander Onderstal<sup>b</sup>

November 22, 2012

In the past, many refinements have been proposed to select equilibria in cheap talk games. Usually, these refinements were motivated by a discussion of how rational agents would reason in some particular cheap talk games. In this paper, we propose a behavioral refinement and stability measure that is meant to predict actual behavior in a wide range of cheap talk games. According to our Average Credible Deviation Criterion (ACDC), the stability of an equilibrium is determined by the frequency and size of credible deviations. ACDC organizes the results from cheap talk experiments well, even in cases where other criteria remain silent.

KEYWORDS: cheap talk, neologism proofness, credible deviation, refinement, ACDC, experiment

---

\* We are grateful to Navin Kartik, Stefano Dellavigna and Charlie Holt for insightful advice and Joel Sobel for extensive comments. In addition, we would like to thank seminar participants at the Universities of Aachen, Amsterdam, Bonn, Cologne, East Anglia, and at LMU Munich and New York University for helpful comments. Financial support from the Dutch National Science Foundation (NWO-VICI 453-03-606) and from the University of Amsterdam Research Priority Area in Behavioral Economics is gratefully acknowledged.

<sup>a</sup> Department of Economics, Institute for Management Research, Radboud University Nijmegen, P.O. Box 9108, 6500 HK Nijmegen, the Netherlands; a.grootruiz@fm.ru.nl.

<sup>b</sup> Faculty of Economics and Business, University of Amsterdam, Roetersstraat 11, 1018 WB, the Netherlands; T.J.S.Offerman@uva.nl, onderstal@uva.nl.

# 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.

Several refinements have been proposed to select an equilibrium in cheap talk games. Often, such refinements were based on an intuitive notion of how rational players would reason in the context of a particular set of cheap talk games.<sup>1</sup> For instance, Farrell (1993) and Matthews, Okuno-Fujiwara & Postlewaite (1991) formulated refinements in which equilibria are discarded if they allow senders to submit credible deviating messages.<sup>2</sup> Unfortunately, both Farrell's neologism proofness and Matthews et al.'s (strong) announcement proofness criteria eliminate all equilibria in many games, including the original Crawford-Sobel game.<sup>3</sup> 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 select a plausible equilibrium in specific settings, but fail to discriminate successfully across a wider range of cheap talk games.

In this paper, we propose a behavioral refinement that is meant to predict actual behavior in a wide range of cheap talk applications. Our Average Credi-

---

<sup>1</sup> For a comprehensive review of Sender-Receiver games, see Sobel (2010).

<sup>2</sup> Standard signaling refinements such as Kohlberg & Merten's strategic stability (1986) have no bite in cheap talk games because messages are costless.

<sup>3</sup> Weak announcement proofness tends to eliminate too few equilibria.

ble Deviation Criterion (ACDC) is based on credible deviations but allows for a continuous instead of a binary stability concept. Its main contribution to the literature is that it works: it makes a prediction in many games and its predictions are validated by experimental evidence.

ACDC takes as a point of departure a theory of credible deviations like credible neologisms proofness (Farrell, 1993) or credible announcements (Matthews, Okuno-Fujiwara & Postlewaite, 1991). These theories stipulate conditions under which a message inducing a deviation from equilibrium is credible and thus upsets the equilibrium. The current approach is to assume that all equilibria that admit credible deviations are equally unstable. ACDC, however, assumes that the 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 think ACDC provides an intuitive solution to the equilibrium selection problem for cheap talk games. Human behavior is messy and seldom completely in (or out of) equilibrium. Still, neologism proofness, announcement proofness, and many of the other cheap talk refinements impose a neat binary distinction between stable and unstable equilibria. Whereas such a binary criterion is appropriate for rational agents, it may unnecessarily lose predictive power when applied to human behavior. ACDC solves this problem in two ways. It is able to select among equilibria in a wide range of games and it provides a continuous stability measure for each equilibrium.

We derive the following results. First, we show that an ACDC equilibrium exists under general conditions. In all applications we have come along, there is a unique ACDC equilibrium. Second and more importantly, the predictions of

ACDC are validated by existing experiments as well as by new experimental data. Wherever experimental evidence exists, the predictions of ACDC are in line with the data: it performs at least as well as other criteria, if they are predictive, and also makes predictions when other criteria are silent. 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 was able to select this equilibrium in the Crawford-Sobel setting (Chen, Kartik & Sobel, 2008). 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 supported by experimental work on (discrete) Crawford-Sobel games (Dickhaut, McCabe & Mukherji (1995), Cai & Wang (2006), Wang, Spezio & Camerer (2010)).

Furthermore, we find that ACDC organizes the main features of the experimental data of the discrete games analyzed by Blume, DeJong, Kim & Sprinkle (2001), originally intended to test the Partial Common Interest criterion. Finally, ACDC is also successful in organizing the results of new experiments. In De Groot Ruiz, Offerman & Onderstal (2012a), we test the predictions of ACDC in a class of veto-threat games introduced in De Groot Ruiz, Offerman & Onderstal (2012b). The data corroborate the predictions of ACDC. The ACDC equilibrium performs better in games where its ACD is smaller. In addition, in each treatment the ACDC equilibrium predicts best, also in games where all other criteria do not select a unique equilibrium.

This paper has the following structure. In section 2, we motivate, define and illustrate ACDC. In section 3, we apply ACDC to the Crawford-Sobel uniform-quadratic model and compare it to other concepts in this framework. In section 4, we discuss the remaining experimental evidence on ACDC. Finally, section 5 concludes.

## 2. Motivation, Definition, Properties, and Applications

### 2.1. Motivation

Applied theorists who analyze strategic information transmission face the following problem if they analyze a new cheap talk game. The model is likely to have several equilibria and so one would like a concept that selects the most plausible equilibrium, tells one how stable that equilibrium is, and is validated by experimental data. However, currently chances are high that one will not find such a concept for the new cheap talk game for two reasons.

First, existing selection criteria tend to select an equilibrium in specific classes of games but not in all relevant applications. Neologism proofness (Farrell, 1993) and announcement proofness (Matthews, Okuno-Fujiwara & Postlewaite, 1991) provide a strong intuition and make meaningful predictions in specific simple discrete games. However, they fail to select an equilibrium in many applied settings, such as that of Crawford and Sobel (1982). In contrast, NITS is very effective in selecting equilibria in the Crawford-Sobel model (Chen, Kartik & Sobel, 2008). The predictions by NITS have been validated by experimental evidence on the Crawford-Sobel model (Dickhaut, McCabe & Mukherji (1995); Cai & Wang (2006); Wang, Spezio & Camerer (2010)). However, in other games, NITS is often not defined, as we will discuss in section 3. PCI has shown to make a prediction which is borne out by experimental data in particular discrete games (Blume, DeJong, Kim & Sprinkle, 2001), but does not select a partition in many continuous settings, such as that of Crawford-Sobel.

Second, even if current concepts select an equilibrium, they do not tell how stable it is, even though experimental evidence suggest there is a considerable degree of variation in the stability of cheap talk equilibria. For instance, experiments on (discrete versions of) the Crawford-Sobel game show that the most informative NITS equilibrium indeed performs best, but that its stability decreases considerably as the bias parameter increases.



Game A in Table 1 illustrates these two issues of selection and stability. In Game A, the Sender sends a costless message  $m$  to the Receiver, who then takes one of three actions:  $a_1, a_2$  or  $a_3$ .<sup>4</sup> The payoffs for both players depend on the Receiver's action and the Sender's type. The Sender's type is private information and is drawn from  $t_1$  and  $t_2$  each with probability  $(1 - \delta) / 2$  and from  $t_3$  with probability  $\delta$ .

TABLE 1  
GAME A

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

*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.  $0 \leq \delta < \frac{1}{2}$  and  $0 \leq \varepsilon < 1$

Game A has two equilibria. We say a type  $t$  induces action  $a$ , if the Receiver always takes action  $a$  after any message  $t$  sends in equilibrium. In the pooling equilibrium, all Senders induce  $a_5$ . In the partially separating equilibrium,  $t_1$  induces  $a_1$ , whereas  $t_2$  and  $t_3$  induce  $a_4$ .

What do credible neologisms (Farrell, 1993) do in this game? Neologisms are out-of-equilibrium messages which are assumed to have a literal meaning in a pre-existing natural language.<sup>5</sup> 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

<sup>4</sup> We will refer to the Sender as a 'she' and the Receiver as 'he.'

<sup>5</sup> See Blume, DeJong, Kim & Sprinkle (1998) and Agranov & Schotter (2012) for studies on the role of natural language in cheap talk games.

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.

If  $\delta = 0$  or  $\varepsilon = 0$ , neologism proofness provides a compelling reason why the partially separating equilibrium is more plausible. The pooling equilibrium admits the credible neologism  $\langle a_4, \{t_2, t_3\} \rangle$ . Hence, it is likely to be unstable as types  $t_2$  and  $\langle a_4, \{t_2, t_3\} \rangle$  can credibly separate themselves from  $t_1$ . On the other hand, the partially separating equilibrium is neologism proof: it admits no credible neologisms and is stable. For  $\delta > 0$  and  $\varepsilon > 0$ , a key limitation of neologism proofness becomes evident. In this case, the partially separating equilibrium also admits a credible neologism, to wit  $\langle a_3, \{t_3\} \rangle$ . This leaves us with no stable equilibrium and no prediction.

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 a concept may lose predictive power.<sup>6</sup>

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

---

<sup>6</sup> We consider equilibrium 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.

equilibrium more frequently than behavior close to the pooling equilibrium. This implies two things. First, it may be possible to select the most plausible equilibrium in a game, even though no equilibrium exists that is entirely stable. Second, to describe behavior in a cheap talk game one needs a continuous stability measure and not just a binary criterion.

## 2.2. Definition and General Results

Our intuition is that, from a behavioral perspective, the 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 the ACDC criterion. We first provide a definition of ACDC and apply it to the Crawford-Sobel game in the following section.

For simplicity, we define ACDC for Sender-Receiver games and pure strategies.<sup>7</sup> Nature draws the Sender type  $t$  from probability density  $f$  on  $T$ , where  $T$  is a compact metric space. The Sender then privately observes her type  $t$  and chooses a costless 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.<sup>8</sup> Let  $U^R : A \times T \rightarrow \mathbb{R}$  be the utility function of the Receiver  $U^S : A \times T \rightarrow \mathbb{R}$  that of the Sender. We assume both are bounded from above and below. A strategy for the Sender consists of a function  $\mu : T \rightarrow M$ , and a strategy of the Receiver is a function  $\alpha : M \rightarrow A$ . Let  $\Sigma^S$  be the set of Sender strategies and  $\Sigma^R$  the set of Receiver strategies. Let  $\{\mu, \alpha\}$  be a strategy profile and  $\Sigma$  the set of all strategy profiles. Finally, let the Receiver have prior beliefs  $\beta^0(t) = f(t)$  and posterior

---

<sup>7</sup> It is more cumbersome but conceptually analogous to formulate ACDC for mixed strategies and/or veto-threat games. In Appendix B, we define ACDC for veto-threat games.

<sup>8</sup> This representation allows for  $T$  and  $A$  to be de facto discrete, by allowing  $U^S$  and  $U^R$  to be constant on regions of the type and outcome space.

beliefs  $\beta(t | m)$  respectively. A pure strategy perfect Bayesian equilibrium (henceforth just ‘equilibrium’)  $\sigma = \{\mu, \alpha, \beta\}$  is characterized by the following three conditions:

$$\begin{aligned} & \text{For each } t \in T, \mu(t) \in \arg \max_{m \in M} U^S(\alpha(m), t) \\ (1) \quad & \text{For each } m \in M, \alpha(m) \in \arg \max_{a \in A} \int_T U^R(a, t) \beta(t | m) dt \\ & \beta(m) \text{ is derived from } \mu \text{ and } \beta^0 \text{ using Bayes Rule whenever possible} \end{aligned}$$

Let  $\Sigma^*$  be the set of equilibria. ACDC provides a stability measure and a selection criterion for equilibria in  $\Sigma^*$ . The starting point of ACDC is a theory of credible deviations  $\gamma$ . Such a theory associates a unique deviating profile  $\gamma(\sigma) = \{\mu^\gamma, \alpha^\gamma\} \in \Sigma$  with an equilibrium  $\sigma$ . 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  $\gamma$ , then  $\gamma(\sigma) = \sigma$ .

We then define the Average Credible Deviation (ACD) of an equilibrium  $\sigma$  relative to  $\gamma$  as:

$$(2) \quad ACD_\gamma(\sigma^*) = E_t[CD_\gamma(t, \sigma^*)]$$

where  $CD_\gamma(t, \sigma)$  measures the *intensity* of type  $t$ 's credible deviation. This captures the likelihood of a particular deviation and the degree to which a deviation upsets the equilibrium. We have selected an intensity measure on the following grounds. It should be

- invariant to affine transformations of payoffs;
- increasing in the difference between the deviating and equilibrium payoff;

Additionally, we think it is desirable that it is decreasing in the lowest rationalizable payoff, as that measures how risky it is to deviate from equilibrium for the Sender. Finally, we think it is convenient to normalize it so that it is 0 if

the difference between deviating and equilibrium payoff is zero and 1 if the difference between deviating and equilibrium payoff is maximal. Specifications of  $CD_\gamma(t, \sigma^*)$  that adhere to the properties above are likely to lead to similar conclusions. We propose the following simple function for  $CD_\gamma(t, \sigma^*)$ . Let  $\Sigma^\dagger$  be the set of rationalizable strategy profiles. Then,  $\underline{U}^S(t) \equiv \inf_{\{\alpha, \mu\} \in \Sigma^\dagger} U^S(t, \alpha(\mu(t)))$  and  $\bar{U}^S(t) \equiv \sup_{\{\alpha, \mu\} \in \Sigma^\dagger} U^S(t, \alpha(\mu(t)))$  are the lowest and highest rationalizable payoff for Sender type  $t$ . Then

$$(3) \quad CD_\gamma(t, \sigma) = \frac{U^S(t, \alpha^\gamma(\mu^\gamma(t))) - U^S(t, \alpha(\mu(t)))}{\bar{U}^S(t) - \underline{U}^S(t)}$$

if  $U^S(t, \alpha(\mu(t))) > \underline{U}^S(t)$ . If  $U^S(t, \alpha(\mu(t))) = \underline{U}^S(t)$ , 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 some form of credible announcements (or, in principle, on any theory of credible deviations). Our preference is to employ credible neologisms.<sup>9</sup> In our view, the simplicity of neologism proofness makes it the most apt to describe the behavior of boundedly rational individuals. We agree with the observations in Matthews, Okuno-Fujiwara & Postlewaite (1991) about the limitations of neologism proofness for rational agents. However, the motivation behind ACDC is to predict behavior and explain experimental data. Hence, our aim is somewhat different from that of most of the credible deviations literature, whose main concern is to establish what would be credible from a rational perspective.

In the case of full rationality, strongly credible announcements are probably the proper basis for a deviation theory, as these are credible even if players anticipate deviations. Credible neologisms assume that players do not anticipate deviations and hence assume some form of bounded rationality. Nonetheless, in

---

<sup>9</sup> We assume that, if a type can send multiple neologisms, she sends the neologism that gives her the highest payoff.

almost all games strongly credible announcements predict that all equilibria are stable, which is clearly not what the empirical data shows. A reason may well lie in the bounded rationality of most people. Estimations of level- $k$  reasoning and cognitive hierarchies of human data in experiments show that in most games the large majority of people engage in level-1 and level-2 reasoning (Nagel, 1995; Camerer, Ho, & Chong, 2004). Cai & Wang (1996) find specifically in a Crawford-Sobel game that the majority of people do not exceed level-2 reasoning.

Determining the credibility of an out-of-equilibrium message also requires quite a bit of strategic sophistication. In analogy to a level- $k$  analysis, one can get a feel for this by looking at how deep people must reason in order to find a particular deviation credible. In a level- $k$  model, players without strategic sophistication ('level-0') will send nor understand deviations. Players with elementary strategic sophistication ('level-1') can understand deviations. However, they will not send them, as they believe they will not be understood. Nor will they anticipate them. Players with a decent amount of strategic sophistication ('level-2') will send credible deviations; but they will not anticipate them. Only sophisticated players ('level-3') will anticipate credible deviations but even they do not expect others to anticipate credible deviations. Very sophisticated players ('level-4') anticipate credible deviations and anticipate that others anticipate them as well. In line with results on strategic reasoning in previous experiments, we expect that credible neologisms and weakly credible announcements (and ACDC with respect to them) come closest to describing the behavior of the large majority of people. Thus, most boundedly rational Receivers will not deviate from equilibrium unless they receive a credible deviation. This means that the assumption implicit in ACDC that an equilibrium is only destabilized if the Sender sends a credible deviation, can be taken as a good approximation of behavior.

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

**Definition 1** *An equilibrium  $\sigma^*$  is an ACDC equilibrium relative to deviation theory  $\gamma$  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  $\sigma$  is an ACDC equilibrium if  $\gamma(\sigma) = \sigma$ .

The following result is immediate.

**Proposition 1** *If the number of equilibrium outcomes is finite, the cheap talk game has an ACDC equilibrium relative to  $\gamma$ .*

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, mild conditions can be formulated in order to guarantee existence of an ACDC equilibrium:

**Proposition 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  $\gamma$ .*

**Proof**  $ACD_\gamma(s)$  achieves a minimum on each compact subset  $S_i$  and thus on  $S$ . Hence,  $\min_{\Sigma^*} ACD_\gamma(\sigma)$  is nonempty and an ACDC equilibrium exists. *Q.E.D.*

Proposition 2 is informative for continuous games for which the equilibrium outcome set is known but not finite. This proposition implies 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.<sup>10</sup> For instance, the class of continuous veto-threats games we introduce in De Groot Ruiz, Onderstal and Offerman (2012b) and for which we derive the ACDC equilibrium in Online Appendix B has an infinite equilibrium outcome set that meets the conditions of Proposition 2.

### 2.3. Applying ACDC

We will now apply ACDC to Game A in Table 1 and to Cai & Wang's (1996) discrete versions of the Crawford-Sobel game. In a discrete game, the ACD of a pure equilibrium  $\sigma$  (with a pure  $\gamma(\sigma)$ ) reduces to

$$\sum_{t \in T} f(t) \frac{U^S(t, \alpha^\gamma(\mu^\gamma(t))) - U^S(t, \alpha(\mu(t)))}{\bar{U}^S(t) - \underline{U}^S(t)},$$

where  $f(t)$  is type  $t$ 's prior probability. Recall that in Game A in the pooling equilibrium, all Senders induce  $a_5$  and  $\langle a_4, \{t_2, t_3\} \rangle$  is the unique credible neologism.

Hence, the ACD of the pooling equilibrium is  $\frac{(1-\delta)(2-1)}{2} \frac{(2-1)}{2-0} + \delta \frac{2-1}{2+\varepsilon-0} = \frac{1}{4} + \delta \frac{2-\varepsilon}{8+4\varepsilon}$ . In the partially separating equilibrium  $t_1$  induces  $a_1$ ,

whereas  $t_2$  and  $t_3$  induce  $a_4$  and  $\langle a_3, \{t_3\} \rangle$  is admitted. Hence, the ACD of the

partially separating equilibrium is  $\delta \frac{(2+\varepsilon-2)}{2+\varepsilon-0} = \frac{\delta\varepsilon}{2+\varepsilon}$ . It is readily verified

that the pooling equilibrium's ACD is greater than the partially separating equilibrium's so that the latter is the ACDC equilibrium. In addition, the ACD

---

<sup>10</sup> An equilibrium of a game with a one-dimensional type and action set is a partition equilibrium if there exists a partition  $t_0 < t_1 < \dots < t_{n-1} < t_n$  of  $T$  such that each type in  $[t_{i-1}, t_i]$  induces action  $a_i$  with  $a_1 < a_2 < \dots < a_n$ . Hence, a partition equilibrium is characterized by a vector  $a = (a_1, \dots, a_n)$  and a partition equilibrium outcome set can be represented by a finite union of subsets of  $\mathbb{R}_1, \dots, \mathbb{R}_n$ .



of the partially separating equilibrium goes to zero if  $\delta$  or  $\varepsilon$  goes to zero. Finally, even though it is ACDC, if  $\delta$  or  $\varepsilon$  become large the partially separating equilibrium becomes less stable because a type can deviate frequently or deviations have a high intensity.

What does ACDC do in the seminal Crawford-Sobel setting? Before the NITS criterion (Chen, Kartik & Sobel, 2008), refining the equilibrium set proved elusive. For example, in the leading uniform-quadratic game, no equilibrium is neologism proof. NITS selects the maximum size equilibrium in the Crawford-Sobel game and experimental evidence on discrete versions of the Crawford-Sobel linear quadratic game supports this prediction (Dickhaut, McCabe & Mukherji, 1995; Cai & Wang, 2006; Wang, Spezio & Camerer, 2010). There is, however, one important experimental finding that remains unexplained in the experimental data. As the bias parameter  $b$  decreases, the maximum size equilibrium becomes more stable.<sup>11</sup>

Consider, for instance, the results on a discrete Crawford-Sobel game by Cai & Wang (2006) depicted in Table 2. Applying ACDC to the discrete Crawford-Sobel games is straightforward. For example, for  $b = 2$ , the ACD of the pooling equilibrium is

$$\frac{1}{5} \left( \frac{U^S(5,5) - U^S(5,7)}{U^S(5,7) - U^S(5,1)} + \frac{U^S(7,5) - U^S(7,7)}{U^S(7,9) - U^S(7,1)} + \frac{U^S(9,5) - U^S(9,7)}{U^S(9,9) - U^S(9,1)} \right) \approx 1.37.$$

In Table 2, we provide the ACD of each equilibrium for the treatments of Cai and Wang (2006). ACDC makes two predictions in line with the experimental data. First, ACDC selects the most informative equilibrium. Second, the most informative equilibrium has a lower ACD and thus becomes more stable as  $b$  becomes smaller and the reverse holds for the pooling equilibrium. Both predictions are intuitive. The most informative equilibrium admits ‘fewer’ or ‘smaller’

---

<sup>11</sup> 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.

credible deviations than the pooling equilibrium for all values of  $b$  in Table 2. This could provide an explanation why it predicts better.

TABLE 2  
ACDC IN BASELINE TREATMENTS CAI & WANG (2006)

$b$ <sup>2</sup>	Pooling Equilibrium <sup>1</sup>			Most Informative Equilibrium			
	Credible Neologisms	Error <sup>3</sup>	ACD	Equilibrium <sup>4</sup>	Credible Neologisms	Error	ACD
0.5	$\langle 1, \{1\} \rangle, \langle 3, \{3\} \rangle, \langle 7, \{7\} \rangle, \langle 9, \{9\} \rangle$	.916	.220	$\{1, \{3\}, \{5\}, \{7\}, \{9\}$		-.084	0
1.2	$\langle 1, \{1\} \rangle, \langle 7, \{5, 7, 9\} \rangle, \langle 8, \{7, 9\} \rangle$	.896	.181	$\{1, 3\}, \{5, 7, 9\}$	$\langle 3, \{3\} \rangle, \langle 8, \{7, 9\} \rangle$	.146	.074
2	$\langle 7, \{5, 7, 9\} \rangle$	.734	.137	$\{1\}, \{3, 5, 7, 9\}$	$\langle 7, \{5, 7, 9\} \rangle, \langle 8, \{7, 9\} \rangle$	.234	.099
4	$\langle 6, \{3, 5, 7, 9\} \rangle$	.391	.101	$\{1, 3, 5, 7, 9\}$	$\langle 6, \{3, 5, 7, 9\} \rangle$	.391	.101

*Notes:* In this Sender-Receiver game payoffs are given by  $U^R(a, t) = 110 - 10 \cdot |t - a|^{7/5}$  and  $U^S(a, t) = 110 - 10 \cdot |t + b - a|^{7/5}$ . The type set is  $\{1, 3, 5, 7, 9\}$ , the action set is  $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ . Each type is equally likely.

<sup>1</sup> In the pooling equilibrium, the Receiver takes action 5 regardless of the message.

<sup>2</sup> The baseline treatments only differ in the size of the bias parameter  $b$ .

<sup>3</sup> As prediction error we take the reported difference between the actual and predicted message-type correlation. Cai and Wang also report other measures as message-action and type-action correlations, which yield a similar picture.

<sup>4</sup> In this column, we show the equilibrium type partition. The Receiver's action from a message coming from a partition element is the average of the types in the partition element.

In addition, as  $b$  increases, the most informative equilibrium admits ‘more’ or ‘larger’ credible deviations. This may explain the fact that the prediction error of the most informative equilibrium appears to become larger as  $b$  increases (and the pooling equilibrium appears to predict less bad). One particular feature of the instability of the most informative equilibrium is that unless it is perfectly separating, there appears to be overcommunication. One explanation for overcommunication could be due to lying averse Senders and/or naïve Receivers (Kartik, Ottaviani, & Squintani, 2007). An additional explanation is that credible neologisms not only destabilize but also lead to more information transmission, as can be seen in Table 2. For instance, if  $b = 4$ , the unique pooling equilibrium predicts no information transmission, but credible neolo-

gisms allow types 3 to 9 to separate themselves from type 1 if equilibrium is reached.

### 3. Crawford-Sobel Game

In this section, we apply ACDC to the leading uniform-quadratic case of Crawford & Sobel's (1982) cheap talk game (henceforth 'CS game'). We compare its predictions to those of existing refinements.

In the 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. Crawford & 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

$$(4) \quad 2n(n-1)b < 1.$$

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

$$(5) \quad t_i^n \equiv \frac{i}{n} - 2bi(n-i).$$

for  $i = 0, \dots, n$  and  $n = 1, \dots, n(b)$ . 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

$$(6) \quad a_i^n = \frac{1}{2}(t_{i-1}^n + t_i^n), \quad i = 1, \dots, n.$$

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  $\bar{\tau}$ . Hence, we can characterize neologisms by  $[\underline{\tau}, \bar{\tau}]$

alone, since the Receiver's best response is  $\tilde{a} = \frac{\underline{\tau} + \bar{\tau}}{2}$ . 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  $[0, \bar{\tau}_0^n)$  where

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

Chen, Kartik & Sobel (2008) show that an equilibrium that fails NITS has a credible neologism of this kind and prove that only the size- $n(b)$  equilibrium satisfies NITS. Hence, the credible neologism  $[0, \bar{\tau}_0^n)$  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  $n \in \{2, \dots, n(b) - 1\}$ , there are  $n - 1$  credible neologisms "in the middle." These take the form  $(\underline{\tau}_i^n, \bar{\tau}_i^n)$  for  $i = 1, \dots, n - 1$ , where  $\underline{\tau}_i^n$  [ $\bar{\tau}_i^n$ ] is indifferent between the equilibrium action  $a_i^n$  [ $a_i^{n+1}$ ] and the neologism action  $\tilde{a}_i^n = (\underline{\tau}_i^n + \bar{\tau}_i^n) / 2$ . We obtain:

$$(7) \quad \begin{aligned} \underline{\tau}_i^n &= \frac{3}{4}a_i^n + \frac{1}{4}a_{i+1}^n - 2b \text{ and} \\ \bar{\tau}_i^n &= \frac{1}{4}a_i^n + \frac{3}{4}a_{i+1}^n - 2b, \end{aligned}$$

$$i = 1, \dots, n - 1.$$

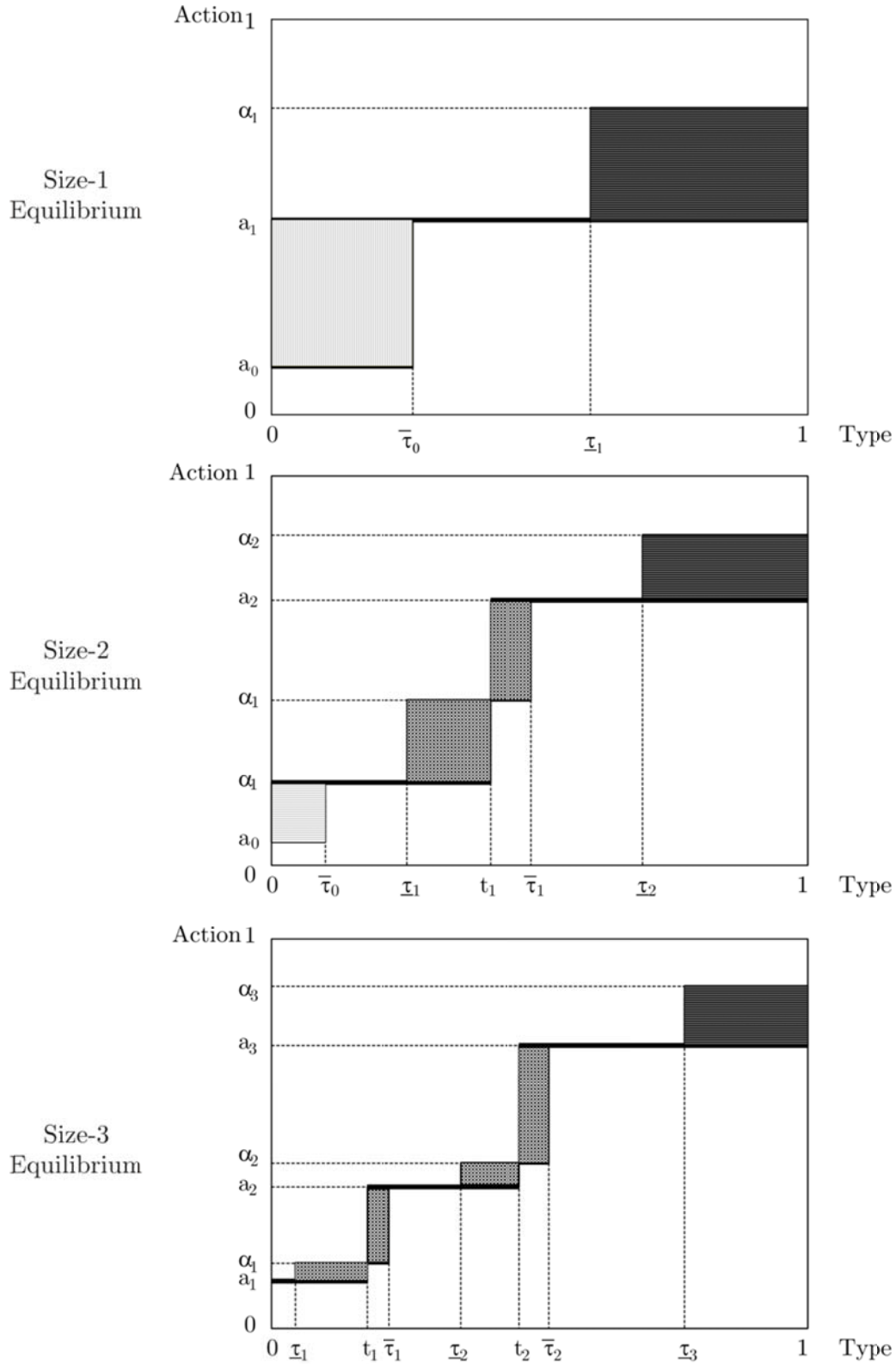


Figure 1

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 their height contributes quadratically to the ACD.

If  $n = n(b)$ , the game has the same types of credible neologisms “in the middle,” with the exception that the neologism  $(\underline{\tau}_1^{n(b)}, \bar{\tau}_1^{n(b)})$  need not exist.<sup>12</sup> Observe that  $\bar{\tau}_{i-1}^n < \underline{\tau}_i^n$  for  $i = 1, \dots, n$ , so that none of the credible neologisms overlap. Figure 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.

**Proposition 3** For all  $b \in \left\{ \frac{1}{10000}, \frac{2}{10000}, \dots, \frac{1}{4} \right\}$  it holds that the ACD of the

size- $n$  equilibrium in the CS game is decreasing in  $n$ .

**Proof** See Appendix A.

**Corollary 1** For all  $b \in \left\{ \frac{1}{10000}, \frac{2}{10000}, \dots, \frac{1}{4} \right\}$ , 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 4** The ACD of the size- $n(b)$  equilibrium tends to zero if  $b$  tends to zero in the CS game.

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

---

<sup>12</sup> If (and only if)  $2bn(b)^2 \geq 1$ , there is no credible neologism of the form  $(\underline{\tau}_1^{n(b)}, \bar{\tau}_1^{n(b)})$  because  $\underline{\tau}_1^{n(b)} = \frac{3}{4}a_1^{n(b)} + \frac{1}{4}a_2^{n(b)} - 2b = -\frac{3}{4n(b)}(2bn(b)^2 - 1) \leq 0$ , which is inconsistent with all types

being in the interval  $[0,1]$  or the interval  $(\underline{\tau}_1^{n(b)}, \bar{\tau}_1^{n(b)}) = (0, t_1^{n(b)})$  being a neologism.

$$\begin{aligned}
\lim_{b \downarrow 0} ACD(\sigma(b)) &\leq \lim_{b \downarrow 0} E_t \left[ \frac{U^S(\tilde{a}^{\sigma(b)}(t), t) - U^S(a^{\sigma(b)}(t), t)}{\min_{t \in T} \{\bar{U}^S(t) - \underline{U}^S(t)\}} \right] \leq \lim_{b \downarrow 0} E_t \left[ \frac{0 - U^S(a^{\sigma(b)}(t), t)}{\frac{1}{4}} \right] \\
&= -4 \cdot \lim_{b \downarrow 0} EU^S \stackrel{1}{=} 4 \cdot \lim_{b \downarrow 0} \left( b^2 + \frac{1}{12n(b)^2} + \frac{b^2(n(b)^2 - 1)}{3} \right) \\
&\leq 4 \cdot \lim_{b \downarrow 0} \left( b^2 + \frac{1}{n(b)^2} + b^2 n(b)^2 \right) \leq 4 \cdot \lim_{b \downarrow 0} \left( b^2 + \frac{4}{(\sqrt{2/b+1} - 1)^2} + \frac{(b + \sqrt{2b + b^2})^2}{4} \right) = 0
\end{aligned}$$

Equality 1 follows from the specification of  $EU^S$  in Crawford & Sobel (1982). Inequality 2 follows from  $n(b) = \left\lfloor \frac{1}{2} + \frac{1}{2} \sqrt{2/b+1} \right\rfloor - 1$  due to (4). The other manipulations are straightforward. *Q.E.D.*

Hence, the ACD of the maximum size 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.

We can now compare ACDC with other criteria. First of all, neologism proofness does not make a prediction: all equilibria admit credible neologisms and are thus unstable. Matthews, Okuno-Fujiwara & Postlewaite (1991) refine neologism proofness with three progressively stronger stability criteria: weak, ordinary and strong announcement proofness. Weak announcement proofness eliminates all equilibria for the same reasons as neologism proofness. Ordinary announcement proofness also tends to eliminate all equilibria and sometimes selects an unintuitive equilibrium. For instance, if  $b \in (\frac{1}{24}, \frac{1}{16})$ , it selects the pooling equilibrium and eliminates the size-2 and size-3 equilibrium.<sup>13</sup> Strong announcement proofness fails to select an equilibrium as it eliminates no equilibrium.

---

<sup>13</sup> or  $b \in (\frac{1}{24}, \frac{1}{16})$ , 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

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. 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 the CS game, 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 type  $t = 1$ ).

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 the CS game, CMR is silent. 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, which cannot be the case under CMR.<sup>14</sup>

The NITS criterion (Chen, Kartik, & Sobel, 2008) is up till now the only refinement based on some notion of stability that can successfully select an equi-

---

types sending the same message  $\{[0, \frac{1}{6} - \frac{4}{6}b], [\frac{2}{6} - \frac{4}{6}b, 1]\}$ . In addition, however, it admits the weakly credible announcement  $\{[0, \frac{1}{5} - \frac{16}{5}b], [\frac{1}{5} - \frac{16}{5}b, \frac{2}{5} - \frac{12}{5}b], [\frac{3}{5} - \frac{12}{5}b, \frac{4}{5} - \frac{16}{5}b], [\frac{4}{5} - \frac{16}{5}b, 1]\}$ . Since for all weakly credible announcements deviating types exist that prefer another weakly credible announcement, none is announcement proof. The size-2 and size-3 equilibria only admit weakly credible announcements composed of the non-overlapping credible neologisms, which are thus credible. Observe that the computational demands on agents to determine whether credible announcements exist and how they look like are quite high.

<sup>14</sup> Rabin also introduces an equilibrium version of CMR, Credible Message Equilibria (CME), but as a consequence of the previous analysis, neither equilibrium in the game can be a CME.



librium in the CS game. 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 general CS game, only the maximum size equilibrium outcome satisfies NITS. The strength of NITS is that it can make predictions under a general monotonicity assumption, and can be justified on the basis of perturbed games with lying averse or non-strategic players.<sup>15</sup>

The prediction of ACDC of Proposition 3 is thus in line with NITS. A difference between NITS and ACDC is that NITS assumes that only the lowest type can separate herself. Hence, according to NITS, the most informative equilibrium is (equally) stable regardless of  $b$ . According to ACDC, also other types can separate through credible deviations. As a consequence, it predicts that the stability of the maximum size equilibrium is decreasing in the bias parameter  $b$ , and if  $b$  is large, the maximum size equilibrium may not be all that stable. The experimental data discussed in the previous section provides support for this prediction.

## 4. Other Experimental Results

Here we discuss the experimental work on equilibrium selection in cheap talk games, in addition to that in the CS game, which we discussed in section 2.3.

### 4.1. 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.

---

<sup>15</sup> A challenge for NITS is that it cannot be applied easily in cheap talk games that have no clear lowest type. For instance, in Game A, the lowest type cannot easily be defined. In section 4.2, we discuss a game that has many NITS equilibria.

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 neologism proofness (and ACDC) are very much aligned, and borne out by the data. In their Game 2 (see Table 3) neologism 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  
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  $\{a_1, \dots, a_5\}$ . 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.

In Blume et al.'s Game 4 (Table 3), no equilibrium is neologism proof while PCI selects a unique equilibrium. 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 neologism 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 neologism 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 neologism where  $t_1$  and  $t_2$  deviate to  $a_4$  and the credible neologism where  $t_3$  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 neologism, 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 neologism 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 with respect to PCI comes from continuous games like the CS game and the veto-threat game which we discuss in the next section. 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.

## 4.2. ACDC in a veto-threat game

In De Groot Ruiz, Onderstal & Offerman (2012a), we test ACDC in fresh experiments. For this, we use games that belong to a class of veto-threat games introduced in De Groot Ruiz, Onderstal & Offerman (2012b). These games are suitable to test ACDC, as they allow for a continuous manipulation of the size and frequency of credible deviations and can have a rich equilibrium set that is

difficult to refine. In Appendix B, we show that ACDC, when adapted for veto-threat games, selects a unique equilibrium in this class of games.

We briefly discuss our results for four treatments (see Table 4). Each treatment is a variation of the following game. The Sender's type is drawn from the uniform distribution over integers in the interval  $[0, B]$ . The Sender sends a costless message  $m \in \{0, 1, \dots, B\}$  to the Receiver, who makes a proposal  $a \in \{0, 1, \dots, B\}$ . The Sender can then accept  $a$  or reject it, in which case the outcome is the disagreement point  $\delta$ . Payoffs for  $a \in \{0, 1, \dots, B\}$  are  $U^R(a) = 60 - \frac{2}{3}a$  and  $U^S(a, t) = 60 - |t - a|$ . In treatments G(120), G(130) and G(210),  $U^R(\delta) = U^S(\delta, t) = 0$ . These treatments only differ in the Boundary parameter  $B$ .

TABLE 4  
FOUR TREATMENTS FROM DE GROOT RUIZ, ONDERSTAL & OFFERMAN (2012A)

Treatment	$U^R(\delta)$	$U^S(\delta)$	$B$	Equilibrium actions <sup>1</sup>	ACD <sup>2</sup>
G(120)	0	0	120	{45}, {0, 60} <sup>**</sup>	0
G(130)	0	0	130	{45}, {0, 60} <sup>*</sup>	0.22
G(210)	0	0	210	{45}, {0, 60} <sup>*</sup>	0.50
T5	0	30	120	{30}, { $a_1, a_1 + 60$ }, {0, $a_2, a_2 + 60$ } <sup>** 3,4</sup>	0

*Notes:* In each game, the Sender sends a message  $m$ , after which the Receiver proposes an action  $a$ . Then the Sender can accept  $a$  or reject  $a$ , in which case the outcome is the disagreement point  $\delta$ .  $t$  was uniformly distributed on the integers in  $[0, B]$ .  $U^R(x) = 60 - \frac{2}{3}x$  and  $U^S(x, t) = 60 - |t - x|$ . <sup>1</sup>An equilibrium has a \* if it is ACDC and \*\* if it is neologism proof as well. <sup>2</sup>The ACD of the ACDC equilibrium. <sup>3</sup> $a_1 \in [0, 30]$  and  $a_2 \in (0, 30]$ . <sup>4</sup>Only {0, 30, 60} is ACDC.

Each of these treatments has a pooling equilibrium where the Receiver always proposes 45 and a partially separating equilibrium, where the Receiver proposes 0 or 60. The only difference is that for  $B = 120$ , the partially separating equilibrium is the unique neologism proof equilibrium, whereas for the other treatments neither equilibrium is neologism proof. For similar reasons as in the CS game, also neologism proofness, the recurrent mop, PCI and CMR do not select an equilibrium. The same holds true for announcement proofness because credible announcements coincide with credible neologisms in this game. Assum-

ing the recurrent mop would converge, neither equilibrium is stable and both are recurrent.<sup>16</sup> CMR can only guarantee that the 0 type can send a credible message (and is silent about what other types do). The only partition that is PCI is  $0 = t_0 < t_1 = B$ .<sup>17</sup> NITS selects the partially separating equilibrium for all  $B$  if one takes 0 as the lowest type.<sup>18</sup>

ACDC selects the partially separating equilibrium in the three treatments and, in addition, predicts that the partially separating becomes less stable as  $B$  increases. In De Groot Ruiz, Onderstal & Offerman (2012a), we find that the data supports the predictions of ACDC. As can be seen in Figure 2, the higher the ACD, the higher the prediction error of an equilibrium. In particular, we find that in each treatment the partially separating equilibrium performs significantly better than the pooling equilibrium. In addition, we find that for  $B = 130$  the partially separating equilibrium performs very similar as when  $B = 120$ , supporting the notion that stability is a continuous characteristic. Finally, we find that the partially separating equilibrium performs significantly better for  $B = 120$  or  $B = 130$  than for  $B = 210$ .

Finally, treatment T5 has  $B=120$ ,  $U^R(\delta) = 0$  and  $U^S(\delta, t) = 30$ . The corresponding game has a continuum of size-2 and size-3 equilibria. None of the earlier refinements selects a unique equilibrium. Even influentiality (selecting the equilibrium with the maximum size) does thus not identify a unique equilibrium

---

<sup>16</sup> The deviation correspondence of the pooling equilibrium (the most interesting case), for instance, contains only message strategies with three messages (say ‘low’, ‘medium’ and ‘high’). In any Receiver strategy in this correspondence, the Receiver proposes 0 after ‘low’, 0 or 45 after ‘medium’ and some higher action after ‘high’; furthermore, the correspondence will contain the strategy in which the Receiver proposes 45 after ‘medium.’ Hence, type  $t = 45$  will separate and send ‘medium’ in any best response to a full-support strategy of the Receiver. Because the deviation correspondence only contains message strategies with three messages, it will not converge to either equilibrium. A similar reasoning holds for the separating equilibrium.

<sup>17</sup> 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  $t = B$ ). 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  $(0, 1]$ .

<sup>18</sup> All types in  $[0, 60]$  are lowest types according to Chen et al.’s definition. The pooling equilibrium survives NITS relative to types in  $[22.5, 105]$ , whereas the separating equilibrium survives NITS relative to types in  $[0, 30]$ .

because there are several size-3 equilibria. Similarly, NITS is not selective as one size-2 and all size-3 equilibria survive NITS. In sum, in De Groot Ruiz, Onderstal & Offerman (2012a), we find that the ACD of an equilibrium predicts well relative to other equilibria and that the ACDC equilibrium has the lowest prediction error.

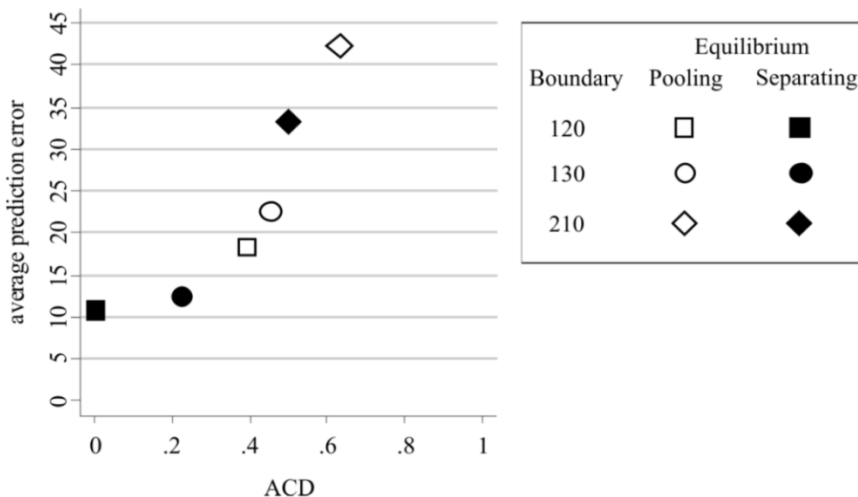


FIGURE 2

The figure plots for each equilibrium in each treatment its theoretical ACD against its empirical prediction error for the last 15 periods. Let  $a^\sigma(t)$  be the equilibrium action of the Receiver given type  $t$  and  $\hat{a}_i(t_i)$  the observed action for observation  $i$ . The average prediction error (for a set of  $n$  observations  $I$ ) is then  $\frac{1}{n} \sum_{i \in I} |\hat{a}_i(t_i) - a^\sigma(t_i)|$ .

$$\frac{1}{n} \sum_{i \in I} |\hat{a}_i(t_i) - a^\sigma(t_i)|$$

## 5. Conclusion

ACDC generalizes refinements based on credible deviations, in particular neologism proofness, capturing the behaviorally relevant aspects of equilibrium stability in cheap talk games. ACDC is based on the intuition that the frequency and size of credible deviations affects equilibrium stability in a continuous rather than a binary manner. ACDC measures the (in)stability of cheap talk equilibria and determines which are most plausible. We showed that an ACDC equilibrium exists under general conditions and that it is unique in a large range of applications. Most importantly, the predictions of ACDC organize the data of

previous and new experiments well. In particular, the data supports the conclusions of ACDC with respect to credible neologisms. Where credible neologisms differ from credible announcements, as in the Crawford-Sobel game, the predictions of ACDC with respect to credible neologisms do better than those with respect to credible announcements.

## References

- Agranov, M., & Schotter, A. (2012). Ignorance is Bliss: an Experimental Study of the Use of Ambiguity and Vagueness in the Coordination Games with Asymmetric Payoffs. *AEJ: Microeconomics*, 4, 77-103.
- Banks, J. S., & Sobel, J. (1987). Equilibrium Selection in Signaling Games. *Econometrica*, 55, 647-661.
- Blume, A., & Sobel, J. (1995). Communication-Proof Equilibria in Cheap-Talk games. *Journal of Economic Theory*, 65, 359-382.
- Blume, A., DeJong, D. V., Kim, Y.-G., & Sprinkle, G. B. (1998). Experimental Evidence on the Evolution of Meaning of Messages in Sender-Receiver Games. *The American Economic Review*, 88, 1323-1340.
- Blume, A., DeJong, D. V., Kim, Y.-G., & Sprinkle, G. B. (2001). Evolution of Communication with Partial Common Interest. *Games and Economic Behavior*, 37, 79-120.
- Blume, A., Kim, Y.-G., & Sobel, J. (1993). Evolutionary Stability in Games of Communication. *Games and Economic Behavior*, 5, 547-575.
- Cai, H., & Wang, J. T. (2006). Overcommunication in Strategic Information Transmission Games. *Games and Economic Behavior*, 56, 7-36.
- Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004). A Cognitive Hierarchy Model of Games. *Quarterly Journal of Economics*, 119, 861-898.
- Chen, Y., Kartik, N., & Sobel, J. (2008). Selecting Cheap Talk Equilibria. *Econometrica*, 76, 117-136.
- Cho, I.-K., & Kreps, D. M. (1987). Signaling Games and Stable Equilibria. *Quarterly Journal of Economics*, 102, 179-221.

- Coles, P., Kushnir, A., & Niederle, M. (Forthcoming). Preference Signaling in Matching Markets. *AEJ: Microeconomics*.
- Crawford, V., & Sobel, J. (1982). Strategic Information Transmission. *Econometrica*, 50, 1431-1451.
- De Groot Ruiz, A. W., Offerman, T., & Onderstal, S. (2012a). An Experimental Study of Credible Deviations and ACDC. *Working paper, University of Amsterdam*.
- De Groot Ruiz, A. W., Offerman, T., & Onderstal, S. (2012b). Power and the Privilege of Clarity: An Analysis of Bargaining Power and Information Transmission. *Working paper, University of Amsterdam*.
- Dickhaut, J. W., McCabe, K. A., & Mukherji, A. (1995). An Experimental Study of Strategic Information Transmission. *Economic Theory*, 6, 389-403.
- Farrell, J. (1993). Meaning and Credibility in Cheap-Talk Games. *Games and Economic Behavior*, 5, 514-531.
- Farrell, J., & Gibbons, R. (1989). Cheap Talk can Matter in Bargaining. *Journal of Economic Theory*, 48, 221-237.
- Gilligan, T. W., & Krehbiel, K. (1990). Organization of Informative Committees by a Rational Legislature. *American Journal of Political Science*, 34, 531-564.
- Kartik, N., Ottaviani, M., & Squintani, F. (2007). Credulity, Lies and Costly Talk. *Journal of Economic Theory*, 134, 93-116.
- Kohlberg, E., & Mertens, J.-F. (1986). On the Strategic Stability of Equilibria. *Econometrica*, 54, 1003-1037.
- Matthews, S. A. (1989). Veto Threats: Rhetoric in a Bargaining Game. *Quarterly Journal of Economics*, 104, 347-369.
- Matthews, S. A., & Postlewaite, A. (1989). Pre-Play Communication in Two-Person Sealed-Bid Double Auctions. *Journal of Economic Theory*, 48, 238-263.
- Matthews, S. A., Okuno-Fujiwara, M., & Postlewaite, A. (1991). Refining Cheap-Talk Equilibria. *Journal of Economic Theory*, 55, 247-273.
- McKelvey, R. D., & Palfrey, T. R. (1998). Quantal Response Equilibrium for Extensive Form Games. *Experimental Economics*, 1, 9-41.



- Morgan, J., & Stocken, P. C. (2003). An Analysis of Stock Recommendations. *The RAND Journal of Economics*, 34, 183-203.
- Morris, S. (2001). Political Correctness. *Journal of Political Economy*, 109, 231-265.
- Myerson, R. B. (1989). Credible Negotiation Statements and Coherent Plans. *Journal of Economic Theory*, 48, 264-303.
- Nagel, R. (1995). Unraveling in Guessing Games: An Experimental Study. *American Economic Review*, 85, 1313-1326.
- Park, I.-U. (1997). Generic Finiteness of Equilibrium Outcome Distributions for Sender-Receiver Cheap-Talk Games. *Journal of Economic Theory*, 76, 431-448.
- Rabin, M. (1990). Communication between Rational Agents. *Journal of Economic Theory*, 51, 144-170.
- Rabin, M., & Sobel, J. (1996). Deviations, Dynamics and Equilibrium Refinements. *Journal of Economic Theory*, 68, 1-25.
- Sobel, J. (2010). Giving and Receiving Advice. *Working paper, University of California, San Diego*.
- Wang, J. T., Spezio, M., & Camerer, C. F. (2010). Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games. *American Economic Review*, 100, 984-1007.

## Appendix A

### Proof of Proposition 3

The proof proceeds as follows. First, we obtain closed-form solutions for the ACD for all  $b$ . Second, we calculate the ACD for the specified values of  $b$ .

The ACD of equilibrium  $\sigma$  in the CS game is equal to

$$\begin{aligned} ACD(\sigma) &= E_t \left[ \frac{U^S(\tilde{a}^\sigma(t), t) - U^S(a^\sigma(t), t)}{\bar{U}^S(t) - \underline{U}^S(t)} \right] \\ &= \int_0^1 \frac{U^S(\tilde{a}^\sigma(t), t) - U^S(a^\sigma(t), t)}{U^S(\min\{t+b, 1\}, t) - \min\{U^S(0, t), U^S(1, t)\}} dt \\ &= \int_0^1 \frac{(a^\sigma - (t+b))^2 - (\tilde{a}^\sigma - (t+b))^2}{-\max\{0, t+b-1\}^2 + \max\{(t+b)^2, (t+b-1)^2\}} dt. \end{aligned}$$

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  $[\underline{t}, \bar{t}]$ . Let  $\hat{t} \equiv \max\{\underline{t}, \min\{\bar{t}, \frac{1}{2} - b\}\}$ . Then,  $\int_{\underline{t}}^{\bar{t}} CD(t, \sigma) dt$  is equal to

$$\begin{aligned} h(b, a^\sigma, \alpha^\sigma, \underline{t}, \bar{t}) &\equiv \int_{\underline{t}}^{\bar{t}} \frac{(a^\sigma - (t+b))^2 - (\tilde{a}^\sigma - (t+b))^2}{\max\{(t+b)^2, (t+b-1)^2\}} dt \\ &= \int_{\underline{t}}^{\hat{t}} \frac{(a^\sigma - (t+b))^2 - (\tilde{a}^\sigma - (t+b))^2}{(t+b-1)^2} dt + \int_{\hat{t}}^{\bar{t}} \frac{(a^\sigma - (t+b))^2 - (\tilde{a}^\sigma - (t+b))^2}{(t+b)^2} dt \\ &= (a^\sigma - \tilde{a}^\sigma) \left\{ \frac{(a^\sigma + \tilde{a}^\sigma - 2)(\hat{t} - \underline{t})}{(b-1+\hat{t})(b-1+\underline{t})} + 2 \log \left[ \frac{b-1+\underline{t}}{b-1+\hat{t}} \right] \right\} \end{aligned}$$

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

As noted in the main text, 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,  $\tilde{a}_i^n, i = 1, \dots, n - 1$ . 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  $\tilde{a}_i^n, i = \underline{i}(b), \dots, n - 1$ , 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  $h(b, a^\sigma, \tilde{a}^\sigma, \underline{t}, \bar{t})$  can be used to calculate the contribution to the ACD for neologisms  $\underline{i}(b) = 1, \dots, n - 1$ . For the highest neologism, the contribution to the ACD is equal to

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

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

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

The ACD of the maximum-size equilibrium is

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

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

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

For each  $b \in \left\{ \frac{1}{10000}, \frac{2}{10000}, \dots, \frac{1}{4} \right\}$ , one can calculate the (closed-form) value of

$ACD(\sigma_b^n)$  for all  $1 \leq n \leq n(b)$ , and verify that the ACD of the size- $n$  equilibrium in the CS game is decreasing in  $n$ .

## Appendix B: ACDC in a Veto-Threat Games

### B.1. Equilibria and ACDC in veto-threat games

Consider the following game. 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. 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  $\delta$ . If  $\delta \in A$ , the game has an internal veto threat and otherwise it has an external veto threat. The outcome set is  $X = A \cup \{\delta\}$ . Let  $U^R : X \times T \rightarrow \mathbb{R}$  be the utility function of the Receiver  $U^S : X \times T \rightarrow \mathbb{R}$  that of the Sender. We assume both are bounded from above and below.

A strategy for the Sender consists of a message strategy  $\mu : T \rightarrow M$  and an acceptance strategy  $\nu : A \times T \rightarrow \{0,1\}$ . The strategy of the Receiver is an action strategy  $\alpha : M^{\text{TM}} A$ . Let  $\Sigma^S$  be the set of Sender strategies and  $\Sigma^R$  the set of Receiver strategies. Let  $\{\mu, \alpha, \nu\}$  be a strategy profile and  $\Sigma$  the set of all strategy profiles. Define  $V^R(x, t; \nu) = U^R(x, t) \cdot \nu(x, t) + U^R(\delta, t) \cdot (1 - \nu(x, t))$  and  $V^S(x, t; \nu) = U^S(x, t) \cdot \nu(x, t) + U^S(\delta, t) \cdot (1 - \nu(x, t))$ . Finally, let the Receiver have prior beliefs  $\beta^0(t) = f(t)$  and posterior beliefs  $\beta(t | m)$  respectively. A pure strategy perfect Bayesian equilibrium (equilibrium henceforth)  $\sigma = \{\mu, \alpha, \beta\}$  is characterized by the following four conditions:

For each  $t \in T$ ,  $m(t) \in \arg \max V^S(\alpha(m), t; \nu)$

$$(8) \quad \text{For each } m \in M, \alpha(m) \in \arg \max_{a \in A} \int_T V^R(a, t; \nu) \beta(t | m) dt$$

$\nu(a, t) = 1$  if  $U^S(a, t) > U^S(\delta, t)$  and  $\nu(a, t) = 0$  if  $U^S(a, t) < U^S(\delta, t)$

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

Let  $\Sigma^*$  be the set of equilibria and  $\Sigma^\dagger$  be the set of rationalizable strategy profiles. Define  $\underline{V}^S(t) \equiv \inf_{\{\alpha, \mu, \nu\} \in \Sigma^\dagger} V^S(t, \alpha(\mu(t)); \nu)$  and  $\bar{V}^S(t) \equiv \sup_{\{\alpha, \mu, \nu\} \in \Sigma^\dagger}$

$V^S(t, \alpha(\mu(t)); \nu)$ . Then

$$(9) \quad CD_\gamma(t, \sigma) = \frac{V^S(t, \alpha^\gamma(\mu^\gamma(t)); \nu^\gamma) - V^S(t, \alpha(\mu(t)); \nu^\gamma)}{\bar{V}^S(t) - \underline{V}^S(t)}$$

if  $V^S(t, \alpha(\mu(t)); \nu^\gamma) > \underline{V}^S(t)$ . If  $V^S(t, \alpha(\mu(t)); \nu) = \underline{V}^S(t)$ . ACDC can be now be defined analogously to the case without a veto by the Sender.

## B.2. ACDC in a veto-threat game

Here we show that ACDC selects a unique equilibrium in the class of veto-threat games introduced by De Groot Ruiz, Offerman & Onderstal (2012b). The games studied experimentally in De Groot Ruiz, Offerman & Onderstal (2012a) belong to this class of games. 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:

$$(10) \quad U^R \text{ on } \mathbb{R} \text{ is twice continuously differentiable, unimodal with a peak at } 0 \text{ and concave.}$$

- (11)  $U^S(x, t)$  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  $f(-z) < y$ ; Finally,  $U^S(\delta) < f(0)$ .<sup>19</sup>

In De Groot Ruiz, Offerman & Onderstal (2012b) we show that only partition equilibria exist. Here we show that there is a unique ACDC equilibrium:

**Proposition 5** *Under assumptions (10) and (11), the unique ACDC equilibrium is the maximum size equilibrium with the highest equilibrium action.*

For the proof of Proposition 5, we introduce some definitions and results from De Groot Ruiz, Offerman & Onderstal (2012b) and derive two helpful lemmas.

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

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

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

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

**Lemma 1** *If  $\langle \tilde{a}, N \rangle$  is a credible neologism relative to equilibrium  $\sigma^*$ , 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

---

<sup>19</sup> Observe that (11) implies assumptions (A2)-(A5) in De Groot Ruiz, Offerman & Onderstal (2012b). Our assumptions here are stricter. In particular, they require a uniform type distribution and a symmetric and concave payoff function for the Sender.

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 (11), 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 2 and 3 in De Groot Ruiz, Offerman & Onderstal (2012b) it follows that in equilibrium

$$(12) \quad \begin{aligned} a_1 &\geq 0, & t_{k-1} - d < a_k \leq t_k - d & \text{ for all } k = 2, \dots, n \text{ and } t_{k-1} + d \leq a_k \text{ for} \\ & & k = 3, \dots, n. \end{aligned}$$

We can now show that under (11), it holds that

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

**Proof.** Due to the  $t$  being uniformly distributed and (11), 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  $[t_k, t_{k+1}]$  ( $k = 0, \dots, n - 1$ ), 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  $a_k \leq t_{k-1} + d$  and by (12), this means  $a_k = t_{k-1} + d \leq t_k - d$  for  $k = 3, \dots, n$ . Now, suppose that  $t_{k-1} + d < t_k - d$  for some  $k = 3, \dots, n$ . 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  $k = 3, \dots, n-1$  is a contradiction with  $a_k \leq t_{k-1} + d$  for  $k = 3, \dots, n$ . Hence,  $a_k = t_{k-1} + d = t_k - d$  for  $k = 3, \dots, n-1$ . Consequently,  $a_k + d = t_k = a_{k+1} - d$  for  $k = 3, \dots, n-1$ .

Furthermore, from the discussion above we have that  $t_2 = a_3 - d$  and that  $a_2 \leq t_1 + d$ . In addition, from (12) it follows that  $a_2 \leq t_2 - d$ . Hence, a necessary condition on  $a_2$  is that  $a_2 \in \arg \max_{t_1+d \leq a \leq t_2-d} (U^R(a) - U^R(\delta))(a + d - t_1)$ . Analogously to the discussion in the proof of Proposition 2 in De Groot Ruiz, Offerman & Onderstal (2012b), 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 5** Suppose that the game has more than one equilibrium outcome. If  $\bar{x} \leq 2d$ , then consider the equilibrium outcome  $\sigma^*$  with  $a_1 = 0$  and  $a_2$  such that  $a_2 \in \arg \max_{a \in \mathbb{R}} U^R(a) (\min\{a + d, 1\} - \frac{1}{2}a_2)$ . If  $\bar{x} > 2d$ , let  $n$  be the natural number for which  $\bar{x} - 2dn \leq 0$  and  $\bar{x} - 2d(n-1) > 0$ , and consider the following  $\sigma^* : a_1 = 0; a_k = \bar{x} - 2d(n-k-2), k = 2, \dots, n$ . We now show that  $\sigma^*$  has the maximum equilibrium size and is the unique ACDC equilibrium outcome.

From Lemma 4 in De Groot Ruiz, Offerman & Onderstal (2012b) and (11), it follows that there exists an  $\bar{x} \in \mathbb{R}$  such that

$$(13) \quad \begin{aligned} U^R(x) - U^R(\delta) + 2dU^{R'}(x) &\geq 0 \text{ for all } x \in [0, \bar{x}] \text{ and} \\ U^R(x) - U^R(\delta) + 2dU^{R'}(x) &< 0 \text{ for all } x \in (\bar{x}, 1-d]. \end{aligned}$$

where a prime ( $'$ ) denotes a derivative with respect to  $x$ . Let  $a^*$  denote the highest equilibrium action  $a_n$  in  $\sigma^*$ . Using (13), it can be verified that  $\sigma^*$  constitutes the highest size equilibrium, analogously to the proof of Proposition

3 in De Groot Ruiz, Offerman & Onderstal (2012b). Similarly, it can be verified that the highest action  $a^{**}$  in any other equilibrium  $\sigma^{**}$  must be smaller than  $a^*$ :

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

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

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

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

To prove that  $\sigma^*$  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  $\sigma^*$  has another credible neologism. By Lemma 1, the set of types that send the credible neologism relative to equilibrium  $\sigma^*$  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 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,  $U^S(a_{k-1}, \tilde{\tau}) < U^S(\tilde{a}, \tilde{\tau}) \leq 0$  and  $U^S(a_k, \tilde{\tau}) < U^S(\tilde{a}, \tilde{\tau}) \leq 0$ . 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  $\sigma^*$ ,  $\sigma^*$  has no other neologisms.

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

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

= 0 implies that

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

From the concavity of  $U^R$  it follows that  $\frac{U^R(a)}{U^{R'}(a)}$  is increasing in  $a$ . Hence,

$\bar{a}^{**} - a^{**}$  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  $\sigma^*$  is an ACDC equilibrium, we show that it has the lowest ACD. By Lemma 2, for equilibrium  $\sigma^{**}$  it must then hold that  $a_1^{**} > 0$  or  $a^{**} < a^*$ . If  $a_1^{**} > 0$ , then a neologism  $\langle \tilde{a}_0, [0, \bar{\tau}_0] \rangle$  exists with  $\tilde{a}_0 < a_1^{**}$ .<sup>20</sup> Suppose now that  $a^{**} < a^*$ . If  $\sigma^*$  does not admit a credible neologism, it is evident that  $ACD(\sigma^*) = 0 < ACD(\sigma^{**})$ . Hence, suppose that  $\sigma^*$  admits the credible neologism  $\langle \tilde{a}^*, [\underline{\tau}^*, 1] \rangle$ .

---

<sup>20</sup> If  $a_1^{**} \geq 2d$ ,  $\tilde{a}_0 = d$  and  $U^S(d, \tilde{\tau}_0) = U^S(a_1^{**}, \tilde{\tau}_0)$ . If  $a_1^{**} \leq d$ ,  $\tilde{a}_0 = 0$  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  $\tilde{\tau}_0 = a_1^{**}$ .

We can now compare the ACD of  $\sigma^*$  and  $\sigma^{**}$ . 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 (11),  $U^S(a, t)$  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  $ACD(\sigma^{**}) = E_t[CD^{\sigma^{**}}(t)] > E_t[CD^{\sigma^*}(t)] = ACD(\sigma^*)$ .

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