

Bargaining over a common conceptual space^{*}

MARCO LICALZI

[licalzi@unive.it]

Dept. of Management

Università Ca' Foscari Venezia

NADIA MAAGLI

[maagli.nadia@gmail.com]

EDE-EM

Université Paris 1-Panthéon-Sorbonne

(September 2013 — Preliminary)

Abstract. Two agents endowed with different individual conceptual spaces are engaged in a dialectic process to reach a common understanding. We model the process as a simple non-cooperative game and demonstrate three results. When the initial disagreement is focused, the bargaining process has a zero-sum structure. When the disagreement is widespread, the zero-sum structure disappears and the unique equilibrium requires a retraction of consensus: two agents who individually agree to associate a region with the same concept end up rebranding it as a different concept. Finally, we document a conversers' dilemma: such equilibrium outcome is Pareto-dominated by a cooperative solution that avoids retraction.

Keywords: cognitive maps, language differences, semantic bargaining, organisational codes, mental models.

Correspondence to:

Marco LiCalzi	Department of Management Università Ca' Foscari Venezia S. Giobbe, Cannaregio 873 30121 Venice, Italy
Phone:	[+39] 041-234-6925
Fax:	[+39] 041-234-7444
E-mail:	licalzi@unive.it
Skype:	mlicalzi

^{*} We are grateful to P. Gärdenfors, Y. Rinott, and M. Warglien, as well as the seminar audience at LUISS Roma for their comments. Financial support under PRIN 20103S5RN3 and STREP 318723 is acknowledged.

1 Introduction

Conceptual spaces have been introduced in Gärdenfors (2000) as an alternative approach to the problem of modelling cognitive representations. In short, a natural concept is associated with a convex region and a conceptual space consists of a collection of convex regions. The underlying geometric structure resonates with early theories of categorisation based on prototypes (Rotsch, 1975; Mervis and Rotsch, 1981) and has recently been given both evolutionary (Jäger, 2007) and game-theoretic foundations (Jäger et al., 2011).

More recently, Warglien and Gärdenfors (2013) have presented an account of semantics as a mapping between individual conceptual spaces. People create meaning by finding ways to map their own individual conceptual spaces to some common ground. A well-known example is the integration of different cultures within an organisation, when different communication codes must blend into a commonly understood language (Wernerfelt, 2004). Warglien and Gärdenfors (2013) rely on the theory of fixed points to demonstrate the possibility of achieving a common conceptual space. Their approach, however, is merely existential and thus offers no insight in the structure of the possible outcomes associated with the creation of a common conceptual space.

This paper addresses the latter issue by analysing a simple non-cooperative game where two agents, each endowed with his own individual conceptual space, bargain over the definition of a common conceptual space. Agents exhibit stubbornness because they are reluctant to give up on their own conceptual space, but they are engaged in a dialectic process that must ultimately lead to a common ground.

We demonstrate two main phenomena, depending on whether the disagreement between agents' individual spaces is focused or widespread. Under focused disagreement, the bargaining process has a zero-sum structure: agents' stubbornness leads to a unique equilibrium where each concedes as little as possible, and the agent who has a larger span of control over the process ends up better off. Under widespread disagreement, the zero-sum structure disappears and each agent confronts a dilemma: holding on to one of his individual concepts weakens his position on another one. At the unique equilibrium, these conflicting pressures force a retraction of consensus: two agents who individually agree on a region being representative of the same concept agree to remould it in order to minimise conflict. Moreover, we uncover a conversers' dilemma: the equilibrium outcome is Pareto-dominated by the Nash bargaining solution without retraction.

2 Model

There are two agents. Each agent $i = 1, 2$ has his own binary conceptual space over the closed unit disk C in \mathbb{R}^2 . (Our results carry through for any convex compact region C in \mathbb{R}^2 , but this specific choice is both elegant and analytically advantageous because C is invariant to rotations.) Conventionally, we label the two concepts L for Left and R for Right and use them accordingly in our figures.

The agents agree on the classification of two antipodal points in D : they both label $l = (-1, 0)$ as L and $r = (0, 1)$ as R , respectively. When the conceptual space represents a

prototypical categorisation, this is implied by the assumption that the two agents agree on the prototypes for L and R , even if the convex regions underlying their conceptual spaces may differ.

The conceptual space of Agent i over C consists of two convex regions L_i and R_i . Dropping subscripts for simplicity, this may look like in Figure 1. Clearly, the representation is fully

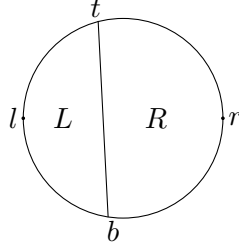


Figure 1: A binary conceptual space.

characterized by the chord \overline{tb} separating the two convex regions. The endpoints t and b for the chord are located in the top and in the bottom semicircumference, respectively. (To avoid trivialities, assume that the antipodal points l and r are interior.) The two regions of the conceptual space may differ in extension and thus the dividing chord need not be a diameter for C .

Consider the conceptual spaces of the two agents. Unless $\overline{t_1 b_1} = \overline{t_2 b_2}$, the regions representing the concepts are different. If the agents are to share a common conceptual space, they must negotiate an agreement. This bargaining process over conceptual spaces amounts to a search for a common ground, where each agent presumably tries to push for preserving as much as possible of his own original conceptual space. Figure 2 provides a pictorial representation for the process: Agent 1 (Primus) and Agent 2 (Secunda) negotiate a shared conceptual space as a compromise between their own conceptual spaces.

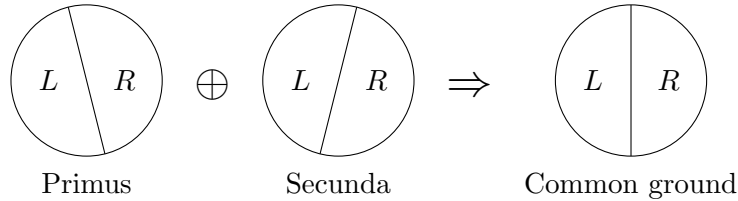


Figure 2: The search for a common conceptual space.

We provide a simple game-theoretic model for their interaction and study the equilibrium outcomes. We do not claim any generality for our model, but its simplicity should help making the robustness of our results transparent.

The two agents play a game with complete information, where the endpoints (t_i, b_i) of each agent i are commonly known. Without any loss of generality, let Primus be the agent for whom t_1 precedes t_2 in the clockwise order. Primus picks a point t in the arc interval $[t_1, t_2]$ from the top semicircumference, while Secunda simultaneously chooses a point b between b_1 and b_2 from the bottom semicircumference. The resulting chord \overline{tb} defines the common

conceptual space. Under our assumption that the antipodal points l and r are interior, the agents cannot pick either of them.

Each agent evaluates the common conceptual space against his own conceptual space. Superimposing these two spaces, there is one region where the common space and the individual space agree on the classification and (possibly) a second region where they disagree. For instance, consider the left-hand side of Figure 3 where the solid and the dotted chords represent the agent's and the common conceptual space, respectively. The two classifications



Figure 3: The disagreement area.

disagree over the central region, coloured in grey on the right-hand side.

Each agent wants to minimise the disagreement between his own individual and the common conceptual space. For simplicity, assume that the payoff for an agent is the opposite of the area of the disagreement region D ; that is, $u_i = -\lambda(D_i)$ where λ is the Lebesgue measure. (Our results carry through for any absolutely continuous measure μ .) Note that the region D need not be convex: when the chords underlying the agent's and the common conceptual space intersect, D consists of two opposing circular sectors.

3 Results

The study of the equilibria is greatly facilitated if we distinguish three cases. First, when $t_1 = t_2$ and $b_1 = b_2$, the two individual conceptual spaces are identical: the unique Nash equilibrium has $t^* = t_1$ and $b^* = b_2$. and the common conceptual space agrees with the individual ones. This is a trivial case, which we consider no further. From now on, we assume that the two individual conceptual spaces disagree; that is, either $t_1 \neq t_2$ or $b_1 \neq b_2$ (or both).

The other two cases depend on the shape of the disagreement region D . When $\overline{t_1 b_1}$ and $\overline{t_2 b_2}$ do not cross inside the disc, then D is a convex set as in the left-hand side of Figure 4. We define this situation as *focused* disagreement, because one agent labels D as L and the



Figure 4: Focused (left) and widespread disagreement (right).

other as R . The disagreement is focused on whether D should be construed as L or R .

Instead, when $\overline{t_1 b_1}$ and $\overline{t_2 b_2}$ cross strictly inside the disc, then D is the union of two circular sectors as in the right-hand side of Figure 4. This is the case of *widespread* disagreement, because the two agents label the two sectors in opposite ways: the top sector is L for one and R for the other, while the opposite holds for the bottom sector.

Focused disagreement. Under focused disagreement, t_1 precedes t_2 and b_2 precedes b_1 in the clockwise order. The disagreement region is convex and the interaction is a game of conflict: as Primus's choice of t moves clockwise, his disagreement region (with respect to the common conceptual space) increases, while Secunda's decreases. In particular, under our simplifying assumption that payoffs are the opposites of the disagreement areas, this is a zero-sum game.

Intuitively, players have opposing interests over giving up on their conceptual spaces. Therefore, we expect that in equilibrium each player concedes as little as possible. In our model, this leads to the stark result that they make no concessions at all over whatever is under their control. That is, they exhibit maximal *stubbornness*. This is made precise in the following theorem, that characterises the unique equilibrium. All proofs are relegated in the appendix.

Theorem 1 *Under focused disagreement, the unique Nash equilibrium is $(t^*, b^*) = (t_1, b_2)$. Moreover, the equilibrium strategies are dominant.*

Figure 5 illustrates the equilibrium outcome corresponding to the situation depicted on the left-hand side of Figure 4. The thick line defines the common conceptual space. In this

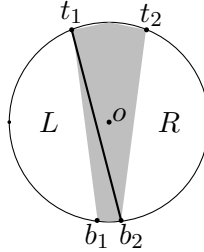


Figure 5: The unique equilibrium outcome under focused disagreement.

example, Primus and Secunda give up the small grey area on the left and on the right of the thick line, respectively. Note how Primus and Secunda stubbornly stick to their own original t_1 and b_1 . Moreover, Primus gives up a smaller area and thus ends up being better off than Secunda. This shows that, in spite of its simplicity, the game is not symmetric. Indeed, the player who controls a larger share of the process leading to a common ground ends up better off than his opponent.

Let $\tau = \widehat{t_1 o t_2}$ (respectively, $\beta = \widehat{b_1 o b_2}$) be the angular distance between t_1 and t_2 (b_1 and b_2). Since Primus chooses a point t in the arc interval $[t_1, t_2]$, the span of his control is τ . Analogously, Secunda picks b in $[b_1, b_2]$ and controls a span β . When $\tau \neq \beta$, the game is

asymmetric and the stronger player is whoever controls the larger span. More precisely, in equilibrium, $u_1(t^*, b^*) - u_2(t^*, b^*)$ has the same sign as $(\tau - \beta)$. Finally, let (t^s, b^s) be the Nash bargaining solution, with t^s and b^s being the midpoints of the respective arc intervals. Then, for $\tau > \beta$, $u_1(t^*, b^*) > u_1(t^s, b^s) = u_2(t^s, b^s) > u_2(t^*, b^*)$, with obvious analogs for $\tau \leq \beta$.

Widespread disagreement. Under widespread disagreement, t_1 precedes t_2 and b_1 precedes b_2 in the clockwise order. The disagreement region is not convex and the interaction is no longer a zero-sum game. We simplify the analysis by making the assumption that the two chords characterising the players' conceptual spaces are diameters. Since this implies $\tau = \beta$, players have the same initial strength and the game is symmetric.

Players' stubbornness now has a double-edged effect, leading to a retraction of consensus at the unique equilibrium. Before stating it formally, we illustrate this result with the help of Fig. 6, drawn for the special case $\tau = \beta = \pi/2$. The thick line depicts the common conceptual

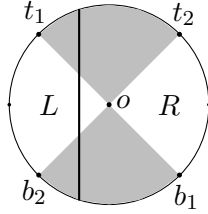


Figure 6: The unique equilibrium outcome under widespread disagreement.

space at the unique equilibrium for this situation.

Consider Primus. Choosing t close to t_1 concedes little on the upper circular sector, but gives up a lot on the lower sector. This temperates Primus' stubbornness and, in equilibrium, leads him to choose a value of t^* away from t_1 . However, as the loss from the lower sector is smaller than the advantage gained in the upper sector, t^* remains closer to t_1 than to t_2 . An analogous argument holds for Secunda.

A surprising side-effect of these tensions is that, in equilibrium, the common conceptual space labels the small white triangle between the thick line and the origin as R, in spite of both agents classifying it as L in their own individual conceptual spaces. That is, in order to find a common ground, players retract their consensus on a small region and switch label. The following theorem characterise the unique equilibrium by means of the two angular distances $\widehat{t^*ot_1}$ and $\widehat{b^*ob_2}$. It is an immediate corollary that the retraction of consensus always occurs, unless $\tau = 0$ and the two agents start off with identical conceptual spaces.

Theorem 2 *Suppose that the individual conceptual spaces are supported by diameters, so that $\tau = \beta$. Under widespread disagreement, there is a unique Nash equilibrium (t^*, b^*) characterised by*

$$\widehat{t^*ot_1} = \widehat{b^*ob_2} = \arctan\left(\frac{\sin \tau}{\sqrt{2} + 1 + \cos \tau}\right).$$

As the equilibrium necessitates a retraction of consensus, it should not be surprising that we have an efficiency loss that we call the *cost of consensus*. The equilibrium strategies lead

to payoffs that are pareto-dominated by those obtained under different strategy profiles. The following result exemplifies the existence of such cost using the natural benchmark provided by the Nash bargaining solution (t^s, b^s) , with t^s and b^s being the midpoints of the respective arc intervals.

Proposition 3 *Suppose that the individual conceptual spaces are supported by diameters. Under widespread disagreement, $u_i(t^*, b^*) \leq u_i(t^s, b^s)$ for each player $i = 1, 2$, with the strict inequality holding unless $\tau = 0$.*

A Proofs

A.1 Proof of Theorem 1

The proof is a bit long, but straightforward. It is convenient to introduce some additional notation. The endpoints (t_i, b_i) for the two agents' chords and their choices for t and b identify six sectors. Proceeding clockwise, these are numbered from 1 to 6 on the left-hand side of Figure 7. For each sector i , we denote its central angle by θ_i ; that is, we let $\theta_1 = \widehat{t_1 o t}$,



Figure 7: Visual aids for the proof of Theorem 1.

$\theta_2 = \widehat{t o t_2}$, $\theta_3 = \widehat{t_2 o b_2}$, $\theta_4 = \widehat{b_2 o b}$, $\theta_5 = \widehat{b o b_1}$, and $\theta_6 = \widehat{b_1 o t_1}$. The following lemma characterises the disagreement area of each player as a function of the six central angles.

Lemma 1 *The disagreement areas for Primus and Secunda are, respectively:*

$$\lambda(D_1) = \frac{\theta_1 + \theta_5 + \sin \theta_6 - \sin(\theta_1 + \theta_5 + \theta_6)}{2}, \quad (1)$$

and

$$\lambda(D_2) = \frac{\theta_2 + \theta_4 + \sin \theta_3 - \sin(\theta_2 + \theta_3 + \theta_4)}{2}.$$

Proof. The disagreement region D_1 for Primus can be decomposed into the two sector-like regions $S_1(t_1 b b_1)$ and $S_2(t_1 t b)$ as shown on the right-hand side of Figure 7. (The figure illustrates a special case, but the formulas hold in general.) We compute the areas $\lambda(S_1)$ and $\lambda(S_2)$, and then add them up to obtain $\lambda(D_1)$.

Consider $S_1(t_1 b b_1)$. It can be decomposed into two regions: the circular segment from b to b_1 with central angle θ_5 , and the triangle $T(t_1 b b_1)$. The area of a circular segment with

central angle θ and radius r is $r^2(\theta - \sin \theta)/2$, which in our case reduces to $(\theta_5 - \sin \theta_5)/2$. Concerning the triangle, the inscribed angle theorem implies that the angle $\widehat{b_1 t_1 b} = \theta_5/2$; hence, by the law of sines, its area can be written as

$$\frac{\overline{t_1 b} \cdot \overline{t_1 b_1} \cdot \sin(\theta_5/2)}{2}. \quad (2)$$

Finally, by elementary trigonometry, $\overline{t_1 b} = 2 \sin[(\theta_5 + \theta_6)/2]$ and $\overline{t_1 b_1} = 2 \sin[(\theta_6)/2]$. Substituting into (2) and adding up the areas of the two regions, we obtain

$$\lambda(S_1) = \frac{\theta_5 - \sin \theta_5}{2} + 2 \sin\left(\frac{\theta_5}{2}\right) \sin\left(\frac{\theta_6}{2}\right) \sin\left(\frac{\theta_5 + \theta_6}{2}\right).$$

By a similar argument, we obtain:

$$\lambda(S_2) = \frac{\theta_1 - \sin \theta_1}{2} + 2 \sin\left(\frac{\theta_1}{2}\right) \sin\left(\frac{\theta_5 + \theta_6}{2}\right) \sin\left(\frac{\theta_1 + \theta_5 + \theta_6}{2}\right).$$

Summing up $\lambda(S_1)$ and $\lambda(S_2)$, we find

$$\begin{aligned} \lambda(D_1) &= \frac{\theta_1 - \sin \theta_1}{2} + \frac{\theta_5 - \sin \theta_5}{2} \\ &\quad + 2 \sin\left(\frac{\theta_5 + \theta_6}{2}\right) \left[\sin\left(\frac{\theta_1}{2}\right) \sin\left(\frac{\theta_1 + \theta_5 + \theta_6}{2}\right) + \sin\left(\frac{\theta_5}{2}\right) \sin\left(\frac{\theta_6}{2}\right) \right]. \end{aligned} \quad (3)$$

After some manipulations shown separately in the following Lemma 2, this expression simplifies to

$$\lambda(D_1) = \frac{\theta_1 + \theta_5 + \sin \theta_6 - \sin(\theta_1 + \theta_5 + \theta_6)}{2}.$$

The derivation of a specular formula for $\lambda(D_2)$ is analogous. □

Lemma 2 *The expression in (3) for $\lambda(D_1)$ can be rewritten as*

$$\lambda(D_1) = \frac{\theta_1 + \theta_5 + \sin \theta_6 - \sin(\theta_1 + \theta_5 + \theta_6)}{2}.$$

Proof. Let $p = \theta_5/2$ and $q = \theta_6/2$. Then

$$\begin{aligned} \lambda(S_1) &= \frac{2p - \sin(2p)}{2} + 2 \sin(p) \sin(q) \sin(p+q) \\ &= \frac{2p - \sin(2p)}{2} + 2 \sin(p+q) [\cos(p-q) - \cos(p+q)] \\ &= \frac{2p - \sin(2p)}{2} + 2 \sin(p+q) \cos(p-q) - \frac{\sin[2(p+q)]}{2} \\ &= \frac{2p - \sin(2p)}{2} + \frac{\sin(2p) + \sin(2q)}{2} - \frac{\sin[2(p+q)]}{2} \\ &= \frac{2p + \sin(2q) - \sin[2(p+q)]}{2} \\ &= \frac{\theta_5 + \sin(\theta_6) - \sin(\theta_5 + \theta_6)}{2}. \end{aligned}$$

An analogous derivation with $p = \theta_1/2$ and $q = (\theta_5 + \theta_6)/2$ leads to

$$\lambda(S_2) = \frac{\theta_1 + \sin(\theta_5 + \theta_6) - \sin[(\theta_1 + \theta_5 + \theta_6)]}{2}.$$

Summing up $\lambda(S_1)$ and $\lambda(S_2)$ we obtain the target formula for $\lambda(D_1)$. \square

Proof of Theorem 1 We compute Primus' best reply function. Given t_1, b_1, t_2, b_2 , and b , Primus would like to choose t in order to minimise $\lambda(D_1)$. Because of the 1–1 mapping between t and θ_1 , we can reformulate this problem as the choice of the optimal angle θ_1 and compute his best reply with respect to θ_1 . Differentiating (1) from Lemma 1, we find

$$\frac{\partial \lambda(D_1)}{\partial \theta_1} = \frac{1 - \cos(\theta_1 + \theta_5 + \theta_6)}{2} > 0$$

for any argument, because $0 < |\theta_1 + \theta_5 + \theta_6| < 2\pi$ under the assumption that l and r are interior. Since $\lambda(D_1)$ is (strictly) increasing in θ_1 , minimising θ_1 by choosing $t = t_1$ is a dominant strategy for Primus. By a similar argument, $b = b_2$ is a dominant strategy for Secunda. Thus, the unique Nash equilibrium (in dominant strategies) is $(t^*, b^*) = (t_1, b_2)$. \square

A.2 Proof of Theorem 2

Similarly to the above (except for switching b_1 and b_2), the endpoints (t_i, b_i) for the two agents' chords and their choices for t and b identify six sectors. Proceeding clockwise, these are numbered from 1 to 6 on the left-hand side of Figure 8.



Figure 8: Visual aids for the proof of Theorem 2.

For each sector i , we denote its central angle by θ_i . The notation is similar, except that now $\theta_3 = \widehat{t_2 o b_1}$, $\theta_4 = \widehat{b_1 o b}$, $\theta_5 = \widehat{b o b_2}$, and $\theta_6 = \widehat{b_2 o t_1}$. Recall that $\tau = \theta_1 + \theta_2$ and $\beta = \theta_4 + \theta_5$; moreover, since the conceptual spaces are characterised by diameters, $\tau = \beta$. The following lemma characterises the disagreement area of each player as a function of the six central angles.

Lemma 3 *The disagreement areas for Primus and Secunda are, respectively:*

$$\lambda(D_1) = \frac{\theta_1 - \sin \theta_1}{2} + \frac{\theta_4 - \sin \theta_4}{2} + 2 \cos\left(\frac{\theta_1}{2}\right) \cos\left(\frac{\theta_4}{2}\right) \frac{\sin^2(\theta_1/2) + \sin^2(\theta_4/2)}{\sin(\theta_1/2 + \theta_4/2)}, \quad (4)$$

and

$$\lambda(D_2) = \frac{\theta_2 - \sin \theta_2}{2} + \frac{\theta_5 - \sin \theta_5}{2} + 2 \cos\left(\frac{\theta_2}{2}\right) \cos\left(\frac{\theta_5}{2}\right) \frac{\sin^2(\theta_2/2) + \sin^2(\theta_5/2)}{\sin(\theta_2/2 + \theta_5/2)}.$$

Proof. The disagreement region D_1 for Primus can be decomposed into the two sector-like regions $S_1(t_1tk)$ and $S_2(kb_1b)$ as shown on the right-hand side of Figure 8. We compute the areas $\lambda(S_1)$ and $\lambda(S_2)$, and then add them up to obtain $\lambda(D_1)$.

The region $S_1(t_1tk)$ can be decomposed into two parts: the circular segment from t_1 to t with central angle θ_1 , and the triangle $T(t_1tk)$. The area of the circular segment is $(\theta_1 - \sin \theta_1)/2$. The computation of the area of the triangle needs to take into account that the position of k depends on t . We use the ASA formula: given the length a of one side and the size of its two adjacent angles α and γ , the area is $(a^2 \sin \alpha \sin \gamma) / (2 \sin(\alpha + \gamma))$. We pick $a = \overline{tt_1}$, $\alpha = \widehat{kt_1t}$, and $\gamma = \widehat{t_1tk}$. By the inscribed angle theorem, $\alpha = (\pi - \theta_1)/2$ and $\gamma = (\pi - \theta_4)/2$. Recall that $\overline{tt_1} = 2 \sin(\theta_1/2)$; moreover, $\sin \alpha = \sin((\pi - \theta_1)/2) = \cos(\theta_1/2)$ and, similarly, $\sin \gamma = \cos(\theta_4/2)$. Hence,

$$\lambda(T) = \frac{2 (\sin(\theta_1/2))^2 \cdot \cos(\theta_1/2) \cdot \cos(\theta_4/2)}{\sin(\theta_1/2 + \theta_4/2)}.$$

Adding up the two areas, we obtain

$$\lambda(S_1) = \frac{\theta_1 - \sin \theta_1}{2} + \frac{2 (\sin(\theta_1/2))^2 \cdot \cos(\theta_1/2) \cdot \cos(\theta_4/2)}{\sin(\theta_1/2 + \theta_4/2)}.$$

By a similar argument,

$$\lambda(S_2) = \frac{\theta_4 - \sin \theta_4}{2} + \frac{2 (\sin(\theta_4/2))^2 \cdot \cos(\theta_1/2) \cdot \cos(\theta_4/2)}{\sin(\theta_1/2 + \theta_4/2)}.$$

Summing up $\lambda(S_1)$ and $\lambda(S_2)$ provides the formula for $\lambda(D_1)$. The derivation of a specular formula for $\lambda(D_2)$ is analogous. \square

A direct study of the sign of the derivative $\partial \lambda(D_1) / \partial \theta_1$ is quite involved, but the following lemma greatly simplifies it. An analogous result holds for Secunda.

Lemma 4 *Let $a = \cos(\theta_4/2)$, $b = \sin(\theta_4/2)$, $c = ab = \sin(\theta_4)/2$, and $x = \tan(\theta_1/4)$. Then*

$$\operatorname{sgn} \left[\frac{\partial \lambda(D_1)}{\partial \theta_1} \right] = \operatorname{sgn} [P(x)], \quad (5)$$

where

$$P(x) = - \left[c (1 + x^2)^2 - 2 (\sqrt{2} + 1) x (1 - x^2) \right].$$

Proof. Differentiating (4) from Lemma 3 and using a few trigonometric identities, we obtain

$$\begin{aligned}\frac{\partial \lambda(D_1)}{\partial \theta_1} &= \frac{1 - \cos \theta_1}{2} + \frac{2 \sin(\theta_1/2) \cos^2(\theta_1/2) \cos(\theta_4/2)}{\sin(\theta_1/2 + \theta_4/2)} - \frac{\cos(\theta_4/2) [\sin^2(\theta_1/2) + \sin^2(\theta_4/2)]}{\sin^2(\theta_1/2 + \theta_4/2)} \\ &\quad \cdot \left[\sin(\theta_1) \sin(\theta_1/2 + \theta_4/2) + \cos(\theta_1) \cos(\theta_1/2 + \theta_4/2) \right] \\ &= \sin^2(\theta_1/2) + \frac{\sin(\theta_1) \cos(\theta_1/2) \cos(\theta_4/2)}{\sin(\theta_1/2 + \theta_4/2)} - \frac{\cos^2(\theta_4/2) [\sin^2(\theta_1/2) + \sin^2(\theta_4/2)]}{\sin^2(\theta_1/2 + \theta_4/2)}\end{aligned}$$

Let $a = \cos(\theta_4/2)$, $b = \sin(\theta_4/2)$, and $x = \tan(\theta_1/4)$. Recall the double angle formulæ $\sin(\theta_1/2) = 2x/(1+x^2)$ and $\cos(\theta_1/2) = (1-x^2)/(1+x^2)$. Then

$$\sin\left(\frac{\theta_1 + \theta_4}{2}\right) = a \left(\frac{2x}{1+x^2}\right) + b \left(\frac{1-x^2}{1+x^2}\right) = \frac{2ax + b(1-x^2)}{1+x^2}.$$

Substituting with respect to the new variable x , we find

$$\begin{aligned}\frac{\partial \lambda(D_1)}{\partial \theta_1} &= \left(\frac{2x}{1+x^2}\right)^2 + \frac{4ax(1-x^2)^2}{(1+x^2)^2 [2ax + b(1-x^2)]} - \frac{a^2 [4x^2 + b^2(1+x^2)^2]}{[2ax + b(1-x^2)]^2} \\ &= -\frac{N(x)}{(1+x^2)^2 [2ax + b(1-x^2)]^2},\end{aligned}\tag{6}$$

where, using the identity $a^2 + b^2 = 1$, the polynomial in the numerator can be written as

$$N(x) = a^2 (1+x^2)^2 [4x^2 + b^2 (1+x^2)^2] - 4ax (1-x^2)^2 [2ax + b(1-x^2)].$$

Let $c = ab = \sin(\theta_4)/2$ and rewrite $N(x)$ after collecting terms with respect to c :

$$\begin{aligned}N(x) &= c^2 (1+x^2)^4 - 4cx (1-x^2) (1+x^2)^2 - 4x^2 (1-x^2) \\ &= \left[c (1+x^2)^2 - 2x (1-x^2) \right]^2 - \left[2\sqrt{2}x (1-x^2) \right]^2 \\ &= \left[c (1+x^2)^2 - 2(\sqrt{2}+1)x (1-x^2) \right] \cdot \left[c (1+x^2)^2 + 2(\sqrt{2}-1)x (1-x^2) \right].\end{aligned}$$

As both θ_1 and θ_4 are in the open interval $(0, \pi)$ by construction, we have $x = \tan(\theta_1/4) > 0$ and $c = \sin(\theta_4)/2 > 0$; hence, the second term in the multiplication is strictly positive. Returning to (6), this implies

$$\operatorname{sgn} \left[\frac{\partial \lambda(D_1)}{\partial \theta_1} \right] = -\operatorname{sgn} [N(x)] = \operatorname{sgn} [P(x)],$$

with $P(x) = -\left[c(1+x^2)^2 - 2(\sqrt{2}+1)x(1-x^2) \right]$, as it was to be shown. \square

It is convenient to work with the central angles subtended by the points on the circumference. Recall that, given t_1, t_2, b_1 , and b_2 , Primus and Secunda simultaneously choose t and b ,

respectively. Then Secunda's choice of b is in a 1-1 mapping with the angle $\theta_5 = \widehat{b_2ob}$, while Primus' choice of t has a similar relation to $\theta_1 = \widehat{t_1ot}$.

The following lemma characterizes Primus' best reply using the central angles θ_1 and θ_5 , rather than the endpoints t and b . As it turns out, such best reply is always unique; hence, with obvious notation, we denote it as the function $\theta_1 = r_1(\theta_5)$. Correspondingly, let $\theta_5 = r_2(\theta_1)$ be the best reply function for Secunda. Finally, recall our assumption that the individual conceptual spaces are supported by diameters: this implies that the two angular distances $\tau = \theta_1 + \theta_2$ and $\beta = \theta_4 + \theta_5$ are equal with $0 \leq \tau = \beta < \pi$; moreover, players' initial positions have the same strength and the game is symmetric.

Lemma 5 *The best reply functions for the two players are*

$$r_1(\theta_5) = \arcsin\left(\frac{\sin(\beta - \theta_5)}{\sqrt{2} + 1}\right) \quad \text{and} \quad r_2(\theta_1) = \arcsin\left(\frac{\sin(\tau - \theta_1)}{\sqrt{2} + 1}\right),$$

with $0 \leq \theta_5 \leq \beta$ and $0 \leq \theta_1 \leq \tau$.

Proof. Consider Primus. (The argument for Secunda is identical.) For any θ_5 in $[0, \beta]$, we search which value of θ_1 in $[0, \tau]$ minimises $\lambda(D_1)$. We distinguish two cases.

First, suppose $\theta_5 = \beta$. Then $\theta_4 = 0$ and $\lambda(D_1) = (\theta_1 + \sin \theta_1)/2$. As this function is increasing in θ_1 , the optimal value is $\theta_1^* = 0$.

Second, suppose $\theta_5 < \beta$. We begin by finding the stationary points of $\lambda(D_1)$. Recall that we let $x = \tan(\theta_1/4)$. By Lemma 4, $\partial\lambda(D_1)/\partial\theta_1 = 0$ if and only if $P(x) = 0$; that is, if and only if

$$c = \frac{2(\sqrt{2} + 1)x(1 - x^2)}{(1 + x^2)^2}.$$

Replacing the double angle formulæ $\sin(\theta_1/2) = 2x/(1 + x^2)$ and $\cos(\theta_1/2) = (1 - x^2)/(1 + x^2)$, we obtain

$$c = (\sqrt{2} + 1) \sin\left(\frac{\theta_1}{2}\right) \cos\left(\frac{\theta_1}{2}\right) = (\sqrt{2} + 1) \frac{\sin \theta_1}{2}.$$

On the other hand, since $c = (\sin \theta_4)/2$ by definition and $\theta_4 + \theta_5 = \beta$, this yields

$$\sin \theta_1 = \frac{\sin \theta_4}{\sqrt{2} + 1} = \frac{\sin(\beta - \theta_5)}{\sqrt{2} + 1}.$$

Since $\theta_5 \in [0, \beta]$, the only solutions to this equation are the supplementary angles θ_1' and $\theta_1'' = \pi - \theta_1'$ with

$$\theta_1' = \arcsin\left(\frac{\sin(\beta - \theta_5)}{\sqrt{2} + 1}\right) < \frac{\pi}{2} < \pi - \theta_1' = \theta_1''.$$

These are the stationary points for $\lambda(D_1)$.

Clearly, $\theta' \geq 0$. We claim that $\theta' < \tau$. If $\pi/2 \leq \tau$, this is obvious. Suppose instead $\tau < \pi/2$. Since $\theta_4 < \beta = \tau < \pi/2$, we have $\sin \theta_1' = (\sqrt{2} - 1) \sin(\theta_4) < \sin \theta_4 < \sin \tau$ and thus $\theta_1' < \tau$. We conclude that the stationary point θ_1' belongs to the interval $[0, \tau]$.

For $\theta_1 = 0$, we have $x = 0$ and $P(x)|_{x=0} = -c = -(\sin \theta_4)/2 < 0$. Therefore, we have by continuity that $P(x)$ changes sign from negative to positive in θ_1' and from positive to

negative in θ_1'' . By Lemma 4, this implies that the only local minimisers for $\lambda(D_1)$ in the compact interval $[0, \tau]$ are $\theta = \theta_1'$ and $\theta = \tau$. Comparing the corresponding values for $\lambda(D_1)$, we find

$$\lambda(D_1)|_{\theta_1=\theta'} < \lambda(D_1)|_{\theta_1=0} < \lambda(D_1)|_{\theta_1=\tau},$$

where the first inequality follows from the (strict) negativity of $\partial\lambda(D_1)/\partial\theta_1$ in $[0, \theta')$ and the second inequality from a direct comparison. Hence, the global minimiser is θ' . Combining the two cases, it follows that, for any θ_5 in $[0, \beta]$, the unique best reply is $r_1(\theta_5) = \arcsin [\sin(\beta - \theta_5) / (\sqrt{2} + 1)]$. \square

Proof of Theorem 2 A Nash equilibrium is any fixed point (θ_1, θ_5) of the map

$$\begin{pmatrix} \theta_1 \\ \theta_5 \end{pmatrix} = \begin{pmatrix} r_1(\theta_5) \\ r_2(\theta_1) \end{pmatrix}$$

from $[0, \tau] \times [0, \beta]$ into itself. Substituting from Lemma 5 and using $\tau = \beta$, we obtain the system of equations

$$\begin{cases} \sin(\theta_1) = \frac{\sin(\tau - \theta_5)}{\sqrt{2} + 1} \\ \sin(\theta_5) = \frac{\sin(\tau - \theta_1)}{\sqrt{2} + 1} \end{cases} \quad (7)$$

Multiplying across gives

$$\sin(\theta_1) \sin(\tau - \theta_1) = \sin(\theta_5) \sin(\tau - \theta_5);$$

or, using the prosthaphaeresis formula,

$$\cos(2\theta_1 - \tau) - \cos \tau = \cos(2\theta_5 - \tau) - \cos \tau$$

from which we get that the only two possible solutions in $[0, \tau]$ are

$$\theta_1 = \theta_5 \quad \text{or} \quad \theta_1 = \tau - \theta_5.$$

When $\tau > 0$, the second possibility can be discarded because, when replaced in (7), it would yield the contradiction $\theta_1 = \theta_5 = \tau = 0$. (When $\tau = 0$, we trivially obtain $\theta_1 = \theta_2$ as in the first case.) Hence, we are left with $\theta_1 = \theta_5$.

Substituting in the first equation of (7), we obtain

$$\sin(\theta_1) = \frac{\sin(\tau - \theta_1)}{\sqrt{2} + 1} = \frac{\sin \tau \cos \theta_1 - \cos \tau \sin \theta_1}{\sqrt{2} + 1}.$$

As $0 \leq \theta_1 < \pi/2$, dividing by $\cos \theta_1$ yields

$$\tan(\theta_1) = \frac{\sin \tau}{\sqrt{2} + 1 + \cos \tau}$$

and the result follows.

A.3 Proof of Proposition 3

Proof. Recall that the payoff for an agent is the opposite of the area of the disagreement region. Consider Primus. (The proof for Secunda is analogous.) Let D^* and D^s be the region of disagreement between Primus' and the common conceptual space at the equilibrium and, respectively, at the Nash cooperative solution. For $\tau = 0$, $D^* = D^s$. Hence, we assume $\tau \neq 0$ and show that $\lambda(D^s) - \lambda(D^*) > 0$.

At the Nash bargaining solution, $\theta_1^s = \theta_2^s = \tau/2$; replacing these into (5), we find $\lambda(D^s) = \tau/2$. At the Nash equilibrium, $\theta_1^* = \theta_5^*$ and thus $\theta_4^* = \tau - \theta_1^*$; substituting these into (5) and dropping superscripts and subscripts for simplicity, we obtain

$$\lambda(D^*) = \frac{\tau}{2} - \left[\frac{\sin \theta + \sin(\tau - \theta)}{2} \right] + 2 \cos\left(\frac{\theta}{2}\right) \cos\left(\frac{\tau - \theta}{2}\right) \frac{\sin^2(\theta/2) + \sin^2((\tau - \theta)/2)}{\sin(\tau/2)}.$$

Hence, using standard trigonometric identities,

$$\begin{aligned} \lambda(D^s) - \lambda(D^*) &= \left[\frac{\sin \theta + \sin(\tau - \theta)}{2} \right] \\ &\quad - 2 \cos\left(\frac{\theta}{2}\right) \cos\left(\frac{\tau - \theta}{2}\right) \frac{\sin^2(\theta/2) + \sin^2((\tau - \theta)/2)}{\sin(\tau/2)} \\ &= \sin\left(\frac{\tau}{2}\right) \cos\left(\theta - \frac{\tau}{2}\right) \\ &\quad - \left[\frac{1}{\sin(\tau/2)} \right] \left[\cos\left(\frac{\tau}{2}\right) + \cos\left(\theta - \frac{\tau}{2}\right) \right] \left[1 - \frac{\cos \theta}{2} - \frac{\cos(\tau - \theta)}{2} \right] \\ &= \left[\frac{1}{\sin(\tau/2)} \right] \left[1 - \cos^2\left(\frac{\tau}{2}\right) \right] \cos\left(\theta - \frac{\tau}{2}\right) \\ &\quad - \left[\frac{1}{\sin(\tau/2)} \right] \left[\cos\left(\frac{\tau}{2}\right) + \cos\left(\theta - \frac{\tau}{2}\right) \right] \left[1 - \cos\left(\frac{\tau}{2}\right) \cos\left(\theta - \frac{\tau}{2}\right) \right] \\ &= \left[\frac{\cos(\tau/2)}{\sin(\tau/2)} \right] \sin^2\left(\theta - \frac{\tau}{2}\right), \end{aligned}$$

from which we obtain

$$\text{sgn}[\lambda(D^s) - \lambda(D^*)] = \text{sgn}[\tan(\tau/2)].$$

Since $0 < \tau < \pi$, $\tan(\tau/2) > 0$, and the claim follows. \square

References

- [1] R. Clark (2012), *Meaningful Games: Exploring Language with Game Theory*, The MIT Press.
- [2] J. Crémer, L. Garicano and A. Prat (2007), “Language and the theory of the firm”, *Quarterly Journal of Economics* **122**, 373–407.
- [3] R. Dietz (2013), “Comparative concepts”, *Synthese* **190**, 139–170.

- [4] P. Gärdenfors (2000), *Conceptual Spaces: The Geometry of Thought*, The MIT Press.
- [5] G. Jäger (2007), “The evolution of convex categories”, *Linguistics and Philosophy* **30**, 551–564.
- [6] G. Jäger, L.P. Metzger, and F. Riedel (2011), “Voronoi languages: Equilibria in cheap-talk games with high-dimensional types and few signals”, *Games and Economic Behavior* **73**, 517–537.
- [7] C. Mervis and E. Rotsch (1981), “Categorization of natural objects”, *Annual Review of Psychology* **32**, 89–115.
- [8] E. Rotsch (1975), “Cognitive representations of semantic categories”, *Journal of Experimental Psychology: General* **104**, 192–233.
- [9] R. Selten and M. Warglien (2007), “The emergence of simple languages in an experimental coordination game”, *Proceedings of the National Academy of Sciences* **104**, 7361–7366.
- [10] B. Skyrms (2010), *Signals: Evolution, Learning, and Information*, Oxford University Press.
- [11] M. Warglien and P. Gärdenfors (2013), “Semantics, conceptual spaces, and the meeting of minds”, *Synthese* **190**, 2165–2193.
- [12] B. Wernerfelt (2004), “Organizational languages”, *Journal of Economics and Management Strategy* **13**, 461–472.