

Sticks and Carrots: Two Incentive Mechanisms Supporting Intra-Group Cooperation

Thorsten Janus & Jamus Jerome Lim*

April 21, 2008

Abstract

In this note, we introduce two distinct incentive mechanisms that support dynamic intra-group cooperation in the context of prisoner's dilemma payoffs. The first mechanism involves a reward for cooperating, where the rewarding party may be outside a given relationship. The second mechanism involves a punishment for defection, where the punishing party may be outside the relationship. We also discuss how these mechanisms are relevant in real-world groups such as criminal gangs and military platoons.

KEYWORDS: Intra-group cooperation, prisoner's dilemma

JEL CLASSIFICATION: D71, D74

*University of Wyoming and the World Bank, respectively. Emails: tjanus@uwyo.edu and jlim@worldbank.org. The authors thank Donald Wittman for helpful comments and suggestions. Financial support from the Graduate Division, University of California, Santa Cruz (Janus) and the Institute on Global Conflict and Cooperation (Lim) are gratefully acknowledged. The standard disclaimers apply.

1 Introduction

Don Corleone: What is the interest for my family?

Sollozzo: Thirty percent. In the first year your end should be 3, 4 million dollars and then it would go up.

Don Corleone: And what is the interest for the Tattaglia family?

Sollozzo: My compliments. I'll take care of the Tattaglia's, out of my share.

...

Don Corleone: Consigliere of mine, I think it's time you told your Don what everyone seems to know.

Tom Hagen: I didn't tell Mama anything. I was just about to come up and wake you so that I could tell you.

Don Corleone: But you needed a drink first. And now you've had your drink.

Tom Hagen: They shot Sonny on the causeway. He's dead.

The Godfather (1972) (Mario Puzo)

Ever since Hardin (1968) popularized the story giving rise to the prisoner's dilemma scenario, social scientists have sought to find solutions to sustaining cooperative outcomes given the particular payoff structure associated with the dilemma. Theory has shown that cooperation may result because of informational imperfections about players' options, motivations, or behavior (Kreps, Milgrom, Roberts & Wilson 1982); if players are possess sufficiently high (Fudenberg & Maskin 1990) or sufficiently homogeneous (Haag & Lagunoff 2007); if punishments are contagious (Ellison 1994); if principals exist that can randomly delay the arrival of payoffs (Chakravorti, Conley & Taub 1996); or through reciprocity as an evolutionarily stable strategy (Axelrod & Hamilton 1981).

This theoretical focus is justified, for such cooperative outcomes are more than theoretical curiosities: There is a body of experimental evidence that suggests that human subjects do in fact engage in cooperation in the context of a prisoner's dilemma (Andreoni & Miller 1993), and even computer simulations allow for such a possibility (Howard 1988).

Olson (1965) was among the first to formally pose the puzzle of group formation and cooperation, and this has provoked a large literature seeking to understand group behavior. In this note, we introduce two incentive mechanisms to sustain intra-group cooperation with prisoner's dilemma payoffs. For simplicity, we examine three-agent groups where relations may either be triadic—one person interacting with two others—or tripartite, where all agents interact. Due to shirking incentives, sustained group cooperation requires a system of endogenous enforcement. We posit a reward structure based on a transfer system and a punishment structure based on collective punishment. Both can ensure cooperation.

There is an existing literature that seeks to model institutions and social networks in terms of endogenous enforcement. The use of incentive slackness in triadic relations to tie strategies across two party games or "domains," has been studied by Aoki (2001); Bernheim & Whinston (1990) while exogenous superior information or enforcement capability among group members compared to non-group members is used in (Fearon & Laitin 1996; Ghatak & Guinnane 1999). Moreover, such an institutional arrangement may itself be endogenous (Okada

1993). The model here takes the former approach, but we add to the literature in two ways. First, we use transfers to endogenize the amount of incentive slack. Second, our punishment mechanism extends the use of incentive slack to three-party settings.

Tripartite cooperation has also been examined in Hart & Kurz (1983) and Ray & Vohra (1999). While both papers seek to endogenize the process of coalition formation, the latter goes a step further in endogenizing the coalition structure. Our note differs from these papers in that we do not assume that binding contracts can be written. An application in this vein, where contracts are assumed to bind, is the trade bloc model of Burbidge, DePater, Myers & Sengupta (1997).

2 Analytical Framework

The environment is comprised of a group with three agents, which are represented by the set $A = \{1, 2, 3\}$. Each individual i possesses a strategy set S_i , with $S' = S_1 \times S_2 \times S_3$. Let $q_i \in S_i$ be a feasible action for player i in the stage game, and denote $q = \{q_i, q_j, q_k\}$. Choices are perfectly observable, and players have perfect recall. A pure strategy for a player i is thus a sequence $\{s_{i,t}(\cdot)\}_{t=1}^{\infty}$ mapping the history \mathfrak{H}_{t-1} of previous action choices to the action choice in period t , $s_{i,t}(\mathfrak{H}_{t-1}) \in S_i$; with the set of all such pure strategies given by Σ_i . These pure strategy profiles induce an outcome path $Q(s) = \{q_t\}_{t=1}^{\infty} = \{q_{1,t}, q_{2,t}, q_{3,t}\}_{t=1}^{\infty}$.

These agents interact in bilateral relationships over an infinite horizon (or, alternatively, over a finite horizon with no known termination time). These relationships are summarized in Figure 1.

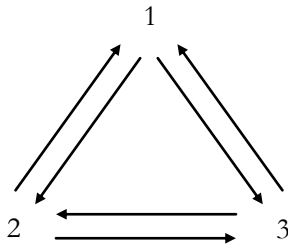


Figure 1: Agent relationships.

Agents have individual welfare given by

$$V_{i,0} = \sum_{t=0}^{\infty} (\delta_i)^t v_{ij,t}(q_{i,t}, q_{j,t}, q_{k,t}; \beta_{i,jk,t}), \quad (1)$$

where $V_{i,0}$ is the date 0 payoff to agent i given the agent's own action $q_{i,t}$ and those of the other players $q_{j,t}$ and $q_{k,t}$, which result from the strategies $s_{i,t}$, $s_{j,t}$, and $s_{k,t}$ employed at time t , respectively; $0 < \delta_i < 1$ is i 's subjective

discount rate; and $\beta_{i,jk,t}$ is a measure of the external benefit accruing to i for a relationship between individuals j and k at time t .

In the prisoner's dilemma, let $s_{i,t} = \{c, d; c, d; \tau_{ij}; \tau_{ik}\}$, where $\tau_{ij} (\tau_{ik}) \geq 0$ is a transfer from i to j (i to k). $v_{ij}(d, c) > v_{ij}(c, c) > v_{ij}(d, d) > v_{ij}(c, d)$ and similarly for the state game between i and k . Therefore, strategy d (c) is defection (cooperation) and the payoff structure does not depend on time.

Assumption 1 (Nash reversion strategy). Agents employ a *Nash reversion strategy* in the group formation game for the infinite horizon. In other words, a player i plays a strategy s_i along the equilibrium path $Q(s)$ until one of the three players defects, and the Nash equilibrium of the stage game $q_{i,t} = \{d; d; \tau_{ij} = 0; \tau_{ik} = 0\}$ is played thereafter.

This trigger strategy is fairly standard in the literature, but here we assume that if even if just one player deviates, all players—even those who have not experienced deviation directly in the prisoner's dilemma—revert to the static Nash equilibrium. In addition, we have included the cessation of transfers as well as the corresponding flows of external benefits.

There is experimental evidence that Nash reversion strategies are employed in repeated game settings, including the prisoner's dilemma, for two person games at least (Engle-Warnick & Slonim 2004; Selten & Stoecker 1986). Nash reversion also allows us to show equilibrium existence for the strongest threat possible; if a cooperative equilibrium does exist, one can always look for lighter punishments to support it. We will say that group formation is sustainable if all agents cooperate pairwise in their prisoner's dilemmas.

Definition 1. A subgame perfect pure strategy Nash equilibrium in the three-player game is a triple $\{q_1^*, q_2^*, q_3^*\}$ induced by a profile $s = (s_1^*, s_2^*, s_3^*) \in \Sigma_1 \times \Sigma_2 \times \Sigma_3$ such that for every h_{t-1} the restriction $s|_{h_{t-1}}$ to the subgame starting at t satisfies: $\forall i, j \in A, i \neq j : \{\nexists s'_i \neq s_i^* \text{ such that } V_i(s'_i|_{h_{t-1}}, s_j^*|_{h_{t-1}}, s_k^*|_{h_{t-1}}) > V_i^*(s_i^*|_{h_{t-1}}, s_j^*|_{h_{t-1}}, s_k^*|_{h_{t-1}})\}$.

We now introduce the central proposition of this note.

Proposition 1 (Group formation). *For $0 < \delta_i, \delta_j, \delta_k < 1$ and $A = \{1, 2, 3\}$, there exist Nash reversion strategies involving rewards and punishments such that group formation is sustainable even when one or more incentive compatibility constraints in the bilateral prisoner's dilemma are violated.*

Proof. See appendix. □

Remark. The result for group formation above is relatively comprehensive. That is, if the conditions in Proposition 1 are satisfied such that the equilibrium in Definition 1 exists, then incentive constraints can even be violated for *both* players in the same prisoner's dilemma.

The proposition thus allows us to assert that, with appropriate Nash reversion strategies, group formation in the context of a game with bilateral prisoner's dilemma payoffs can be sustained even if any two agents meeting alone would

not cooperate. This extra cooperation is sustained either due to a transfer rewards system—carrots—or collective punishment threats—sticks.

For example, if agent 3 enjoys an external benefit from cooperation between 1 and 2, he or she can make 1 internalize this by paying 1 a transfer in every period as long as 1 cooperates with 2. If cooperation ceases, so does the transfer. This is essentially a repeated contingent use of the Coase theorem. Alternatively, though 2 and 3 may both be tempted to defect on each other, the threat that—if that were to occur—1 will defect on both of them in a collective punishment (and hence remove their surplus from their interactions with 1), can remove the temptation.

3 Applications

We offer two applications: Rewards in criminal gangs, and discipline within military platoons.

Given the high risks involved in gang participation, the finding that street-level members of gangs often earn more or less the minimum wage presents a puzzle (Levitt & Venkatesh 2000). Even after taking into account the discounted value of potential future earnings—given the tournament structure of gang participation—membership in a gang appears to be suboptimal, especially given the likelihood that criminal discount rates are relatively high (Paternoster & Brame 1998).

However, the enforcement of the cooperative outcome in gangs appears to be bolstered by rewards offered for cooperative behavior. These rewards may be monetary (a larger share of crime-related profits), nonpecuniary (enhanced recognition from promotion up the gang’s hierarchy, or a greater sense of belonging), or both. In addition, these rewards usually accompany circumstances where group cohesion may be threatened. For example, Klein (1995) discusses how members of street gangs enjoy enhanced cohesion after engaging in sporadic criminal or violent activity, while Decker & van Winkle (1996) document how proceeds from the sale of drugs are considered rewards that are a part of street culture. These can be viewed as *ex post* transfers to members, offered by the gang leader, in exchange for the external benefits—usually viewed as leaders’ prestige—that he or she obtains as a result of the gang’s continued existence.

Punishment in the military is often meted out to the entire group. Ostensibly, this is to build a sense of group “intermindedness,” or more precisely, a sense of unity through social solidarity (Spindler 1948). This is especially cogent in boot camp—where recruits are transitioning from civilian to military life—and in paramilitary organizations (Archer 1999), where looser institutional structures make it harder to prevent desertion.

Such punishment allows military organizations to support group cooperation, since punishments for deviant behavior in the platoon target the collective for the transgressions of any one soldier. For example, mass punishments remain common in boot camps even though the practice is expressly prohibited by military law (Heckathorn 1988). In terms of our model, collective punishments

of this form ensure that compliance rapidly emerges in the platoon.

References

- Andreoni, James A. & John H. Miller (1993). "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence". *Economic Journal* 103(418) (May): 570–585
- Aoki, Masahiko (2001). *Toward a Comparative Institutional Analysis*. Cambridge, MA: MIT Press
- Archer, David (1999). "Exploring "Bullying" Culture in the Para-Military Organisation". *International Journal of Manpower* 20(1/2) (February): 94–105
- Axelrod, Robert & William D. Hamilton (1981). "The Evolution of Cooperation". *Science* 211(4489) (March 27): 1390–6
- Bernheim, B. Douglas & Michael D. Whinston (1990). "Multimarket Contact and Collusive Behavior". *RAND Journal of Economics* 21(1) (Spring): 1–26
- Burbridge, John B., James A. DePater, Gordon M. Myers & Abhijit Sengupta (1997). "A Coalition-Formation Approach to Equilibrium Federations and Trading Blocs". *American Economic Review* 87(5) (December): 940–956
- Chakravorti, Bhaskar, John P. Conley & Bart Taub (1996). "On Uniquely Implementing Cooperation in the Prisoners' Dilemma". *Economic Theory* 8(2) (June): 347–366
- Decker, Scott H. & Barrick van Winkle (1996). *Life in the Gang : Family, Friends, and Violence*. Cambridge Studies in Criminology. New York, NY: Cambridge University Press
- Ellison, Glenn (1994). "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching". *Review of Economic Studies* 61(3) (July): 567–588
- Engle-Warnick, Jim & Robert L. Slonim (2004). "The Evolution of Strategies in a Repeated Trust Game". *Journal of Economic Behavior & Organization* 55(4) (December): 553–573
- Fearon, James D. & David D. Laitin (1996). "Explaining Interethnic Cooperation". *American Political Science Review* 90(4) (December): 715–35
- Fudenberg, Drew & Eric S. Maskin (1986). "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information". *Econometrica* 54(3) (May): 533–554
- Fudenberg, Drew & Eric Maskin (1990). "Evolution and Cooperation in Noisy Repeated Games". *American Economic Review* 80(2) (May): 274–279
- Ghatak, Maitreesh & Timothy W. Guinnane (1999). "The Economics of Lending with Joint Liability: Theory and Practice". *Journal of Development Economics* 60(1) (October): 195–228
- Haag, Matthew & Roger D. Lagunoff (2007). "On the Size and Structure of Group Cooperation". *Journal of Economic Theory* 135(1) (July): 68–89
- Hardin, Garrett (1968). "The Tragedy of the Commons". *Science* 162(3859) (December 13): 1243–1248
- Hart, Sergiu & Mordecai Kurz (1983). "Endogenous Formation of Coalitions". *Econometrica* 51(4) (July): 1047–1064
- Heckathorn, Douglas D. (1988). "Collective Sanctions and the Creation of Prisoner's Dilemma Norms". *American Journal of Sociology* 94(3) (November): 535–562

- Howard, John V. (1988). “Cooperation in the Prisoner’s Dilemma”. *Theory and Decision* 24(3) (May): 203–213
- Klein, Malcolm W. (1995). *The American Street Gang: Its Nature, Prevalence, and Control*. New York, NY: Oxford University Press
- Kreps, David M., Paul Milgrom, John Roberts & Robert Wilson (1982). “Rational Cooperation in the Finitely Repeated Prisoners’ Dilemma”. *Journal of Economic Theory* 27(2) (August): 245–252
- Levitt, Steven D. & Sudhir Alladi Venkatesh (2000). “An Economic Analysis of a Drug-Selling Gang’s Finances”. *Quarterly Journal of Economics* 115(3) (August): 755–789
- Okada, Akira (1993). “The Possibility of Cooperation in an n-Person Prisoners Dilemma with Institutional Arrangements”. *Public Choice* 77(3) (November): 629–656
- Olson, Mancur (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard Economic Studies. Cambridge, MA: Harvard University Press, revised edition
- Paternoster, Raymond & Robert Brame (1998). “The Structural Similarity of Processes Generating Criminal and Analogous Behaviors”. *Criminology* 36(3) (August): 633–670
- Ray, Debraj & Rajiv Vohra (1999). “A Theory of Endogenous Coalition Structures”. *Games and Economic Behavior* 26(2) (January): 286–336
- Selten, Reinhard & Rolf Stoecker (1986). “End Behavior in Sequences of Finite Prisoner’s Dilemma Supergames”. *Journal of Economic Behavior & Organization* 7(1) (March): 47–70
- Spindler, G. Dearborn (1948). “The Military—A Systematic Analysis”. *Social Forces* 27(1) (October 1948–May 1949): 83–88

Appendix

Proof of Proposition 1. Before proceeding with the proof, it is useful to establish the following lemma.

Lemma 1 (Folk theorem). $\exists \underline{\delta}_i < 1$ such that for any $\delta_i \in [\underline{\delta}, 1] \forall i$, a cooperative outcome path $Q = \{q_{i,t}, q_{k,t}\}_{t=1}^{\infty} = \{c, c\}_{t=1}^{\infty}$ can be sustained with a Nash reversion strategy, where $\underline{\delta}$ is the lowest discount factor that can sustain cooperation, given the payoffs V_{ij} .

Proof. The lemma is a direct result of the Folk Theorem for repeated games applied to the context here; see Fudenberg & Maskin (1986) for a formal treatment. \square

Lemma 2 shows proves the proposition when we allow for both transfers and spillovers from relationships. Lemma 3 shows the special case when there are no spillovers and no transfers, and only the stick mechanism—Nash reversion by all three players following a single defection—sustains cooperation.

Lemma 2. *In the group formation game, there exist Nash reversion strategies with transfers $\tau_{ij} > 0$, $i, j = \{1, 2, 3\}$, $i \neq j$ such that group formation is sustainable, even if one or more incentive compatibility constraints are violated in pairwise games in the absence of transfers, provided $\delta_i > 0$ and $\beta_{i,jk} > 0$ are sufficiently large or $\delta_i > 0$ and the incentive slack of one of the players in a different relationship are sufficiently large.*

Proof. Even with the Nash reversion strategies, not all agents may be willing to play the prisoner's dilemma with both the other agents. However, we can develop the proof for the full cooperation case and then specialize it. Using $\tau_{ij} = -\tau_{ji}$, a nonempty set $\{\tau_{12}, \tau_{13}, \tau_{23}\}$ that satisfies the following conditions are sufficient for a Nash equilibrium:

$$v_{12}(d, c) - v_{12}(c, c) + v_{13}(d, c) - v_{13}(c, c) < \frac{\delta_1}{1-\delta_1} \times \{ [v_{12}(c, c) - v_{12}(d, d)] + [v_{13}(c, c) - v_{13}(d, d)] + (\beta_{1,23} - \tau_{12} - \tau_{13}) \}, \quad (\text{A.1a})$$

$$v_{21}(d, c) - v_{21}(c, c) + v_{23}(d, c) - v_{23}(c, c) < \frac{\delta_2}{1-\delta_2} \times \{ [v_{21}(c, c) - v_{21}(d, d)] + [v_{23}(c, c) - v_{23}(d, d)] + (\beta_{2,13} + \tau_{12} - \tau_{23}) \}, \quad (\text{A.1b})$$

$$v_{31}(d, c) - v_{31}(c, c) + v_{32}(d, c) - v_{32}(c, c) < \frac{\delta_3}{1-\delta_3} \times \{ [v_{31}(c, c) - v_{31}(d, d)] + [v_{32}(c, c) - v_{32}(d, d)] + (\beta_{3,12} + \tau_{13} + \tau_{23}) \}. \quad (\text{A.1c})$$

Without loss of generality, assume that $\beta_{1,23}$ is large and that (A.1a) is satisfied, but (A.1b) and (A.1c) are not. Then there potentially exist values of the transfers τ_{12} and τ_{13} such that the direction of the inequalities in (A.1b) and (A.1c) reverse, while the inequality in (A.1a) remains unchanged. This requires $\hat{\tau}_{12} + \hat{\tau}_{13} < \beta_{1,23}$, where $\hat{\tau}_{ij}$ denotes the minimum required transfer to sustain cooperation between i and j . The bound on 1's willingness to subsidize the game between 2 and 3 depends on the external benefit, $\beta_{1,23}$, he or she enjoys, given cooperation between 2 and 3. Notice that if (A.1b) (if (A.1c)) is also satisfied, then the lower bound of $\beta_{1,23}$ can be lower, since $\hat{\tau}_{23}$ may be positive (negative).

However, not all players may be willing to play the prisoner's dilemma with each other. For example, if 1 and 3 do not play the prisoner's dilemma, then Definition 1 can still be satisfied with $\beta_{2,13} = 0$ so that (A.1) simplifies to

$$v_{12}(d, c) - v_{12}(c, c) < \frac{\delta_1}{1-\delta_1} \{ [v_{12}(c, c) - v_{12}(d, d)] + \beta_{1,23} - \tau_{12} - \tau_{13} \}, \quad (\text{A.2a})$$

$$v_{21}(d, c) - v_{21}(c, c) + v_{23}(d, c) - v_{23}(c, c) < \frac{\delta_2}{1-\delta_2} \{ [v_{21}(c, c) - v_{21}(d, d)] + [v_{23}(c, c) - v_{23}(d, d)] + \tau_{12} - \tau_{23} \}, \quad (\text{A.2b})$$

$$v_{32}(d, c) - v_{32}(c, c) < \frac{\delta_3}{1-\delta_3} \{ [v_{32}(c, c) - v_{32}(d, d)] + \beta_{3,21} + \tau_{13} + \tau_{23} \}. \quad (\text{A.2c})$$

Again without loss of generality, for $\beta_{1,23}$ large, (A.2a) is satisfied and transfers can be used to satisfy the other equations, again up the value of the external

benefit 1 enjoys. If, furthermore, 1 and 2 do not play the prisoner's dilemma so that $v_{12}(\cdot) = v_{21}(\cdot) = 0$, Definition 1 simply requires

$$\beta_{1,23} > \tau_{12} + \tau_{13}, \quad (\text{A.3a})$$

$$v_{23}(d, c) - v_{23}(c, c) < \frac{\delta_2}{1-\delta_2} \{ [v_{23}(c, c) - v_{23}(d, d)] + (\tau_{12} - \tau_{23}) \}, \quad (\text{A.3b})$$

$$v_{32}(d, c) - v_{32}(c, c) < \frac{\delta_3}{1-\delta_3} \{ [v_{32}(c, c) - v_{32}(d, d)] + (\tau_{13} + \tau_{23}) \}, \quad (\text{A.3c})$$

which again is satisfied for sufficiently large $\beta_{1,23}$. The cooperation conditions for all other configurations of agents playing the prisoner's dilemma are analogous. \square

Lemma 3. *For $0 < \delta_i < 1$, $i = \{1, 2, 3\}$, there exist Nash reversion strategies with punishments such that group formation is sustainable, even if one or more incentive compatibility constraints are violated, provided every player has incentive slack in at least one relationship and that $\delta_i > 0$ is sufficiently large.*

Proof. To remove the possibility of the transfer mechanism sustaining cooperation, assume that $\beta_{i,jk} = \tau_{ij} = 0 \forall i, j, k$. If both of an agent's incentive compatibility constraints are violated, by Lemma 1, the agent will not cooperate. Therefore, we consider only cases with at most one of each agent's incentive constraints violated. In this case, group formation could still be possible as a Nash equilibrium. Without loss of generality, let $v_{12}(d, c) - v_{12}(c, c) > \frac{\delta_1}{1-\delta_1} [v_{12}(c, c) - v_{12}(d, d)]$ and $v_{13}(d, c) - v_{13}(c, c) < \frac{\delta_1}{1-\delta_1} [v_{13}(c, c) - v_{13}(d, d)]$. Given Nash reversion group formation requires

$$v_{12}(d, c) - v_{12}(c, c) + v_{13}(d, c) - v_{13}(c, c) < \frac{\delta_1}{1-\delta_1} \{ [v_{12}(c, c) - v_{12}(d, d)] + [v_{13}(c, c) - v_{13}(d, d)] \}, \quad (\text{A.4a})$$

$$v_{21}(d, c) - v_{21}(c, c) + v_{23}(d, c) - v_{23}(c, c) < \frac{\delta_2}{1-\delta_2} \{ [v_{21}(c, c) - v_{21}(d, d)] + [v_{23}(c, c) - v_{23}(d, d)] \}, \quad (\text{A.4b})$$

$$v_{31}(d, c) - v_{31}(c, c) + v_{32}(d, c) - v_{32}(c, c) < \frac{\delta_3}{1-\delta_3} \{ [v_{31}(c, c) - v_{31}(d, d)] + [v_{32}(c, c) - v_{32}(d, d)] \}, \quad (\text{A.4c})$$

and (A.4a) can still be satisfied for $v_{13}(c, c)$ sufficiently large. If both (A.4b) and (A.4c) are satisfied in an analogous manner, then group formation is possible. \square

Taken together, Lemmata 2 and 3 exhaust all possible cases, which concludes the proof of the proposition. \square