ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΜΕΘΟΔΟΛΟΓΙΑΣ, ΙΣΤΟΡΙΑΣ ΚΑΙ ΘΕΩΡΙΑΣ ΤΗΣ ΕΠΙΣΤΗΜΗΣ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ "ΒΑΣΙΚΗ ΚΑΙ ΕΦΑΡΜΟΣΜΕΝΗ ΓΝΩΣΙΑΚΗ ΕΠΙΣΤΗΜΗ"

Η ΑΛΤΡΟΥΙΣΤΙΚΗ ΤΙΜΩΡΙΑ ΚΑΙ Η ΕΠΙΔΡΑΣΗ ΤΗΣ ΣΤΗΝ ΕΝΙΣΧΥΣΗ ΚΑΙ ΤΗΝ ΕΔΡΑΙΩΣΗ ΤΗΣ ΣΥΝΕΡΓΑΣΙΑΣ ΚΑΙ ΤΩΝ ΚΑΝΟΝΩΝ ΚΟΙΝΩΝΙΚΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ

ALTRUISTIC PUNISHMENT AND ITS ROLE IN THE REINFORCEMENT AND ESTABLISHMENT OF COOPERATION AND SOCIAL NORMS

ΠΑΡΑΣΚΕΥΗ ΝΑΛΜΠΑΝΤΗ
ΑΜ: 08Μ17

ΕΠΙΒΛΕΠΟΝΤΕΣ ΚΑΘΗΓΗΤΕΣ:
ΑΡΙΣΤΕΙΔΗΣ ΧΑΤΖΗΣ, ΑΝΑΠΛΗΡΩΤΗΣ ΚΑΘΗΓΗΤΗΣ ΕΚΠΑ
ΕΛΠΙΔΑ ΤΖΑΦΕΣΤΑ, ΑΝΑΠΛΗΡΩΤΡΙΑ ΚΑΘΗΓΗΤΡΙΑ ΕΚΠΑ

Αθήνα, 2012

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΜΕΘΟΔΟΛΟΓΙΑΣ, ΙΣΤΟΡΙΑΣ ΚΑΙ ΘΕΩΡΙΑΣ ΤΗΣ ΕΠΙΣΤΗΜΗΣ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ "ΒΑΣΙΚΗ ΚΑΙ ΕΦΑΡΜΟΣΜΕΝΗ ΓΝΩΣΙΑΚΗ ΕΠΙΣΤΗΜΗ"

Η ΑΛΤΡΟΥΙΣΤΙΚΗ ΤΙΜΩΡΙΑ ΚΑΙ Η ΕΠΙΔΡΑΣΗ ΤΗΣ ΣΤΗΝ ΕΝΙΣΧΥΣΗ ΚΑΙ ΤΗΝ ΕΔΡΑΙΩΣΗ ΤΗΣ ΣΥΝΕΡΓΑΣΙΑΣ ΚΑΙ ΤΩΝ ΚΑΝΟΝΩΝ ΚΟΙΝΩΝΙΚΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ

ALTRUISTIC PUNISHMENT AND ITS ROLE IN THE REINFORCEMENT AND ESTABLISHMENT OF COOPERATION AND SOCIAL NORMS

ΠΑΡΑΣΚΕΥΗ ΝΑΛΜΠΑΝΤΗ
ΑΜ: 08Μ17

ΕΠΙΒΛΕΠΟΝΤΕΣ ΚΑΘΗΓΗΤΕΣ:
ΑΡΙΣΤΕΙΔΗΣ ΧΑΤΖΗΣ, ΑΝΑΠΛΗΡΩΤΗΣ ΚΑΘΗΓΗΤΗΣ ΕΚΠΑ
ΕΛΠΙΔΑ ΤΖΑΦΕΣΤΑ, ΑΝΑΠΛΗΡΩΤΡΙΑ ΚΑΘΗΓΗΤΡΙΑ ΕΚΠΑ

Αθήνα, 2012

Θα ήθελα να ευχαριστήσω ιδιαίτερα τους επιβλέποντες καθηγητές Αριστείδη Χατζή και Ελπίδα Τζαφέστα για την πολύτιμη καθοδήγηση και βοήθειά τους κατά τη συγγραφή αυτής της διπλωματικής εργασίας.

ΠΕΡΙΛΗΨΗ

Το πρόβλημα της συνεργασίας στις ανθρώπινες κοινωνίες αποτελεί ένα κεντρικό ερώτημα των κοινωνικών επιστημών, καθώς το εύρος της αποτελεί ένα μοναδικό φαινόμενο στο ζωικό βασίλειο. Πολλές από τις θεωρίες που έχουν διατυπωθεί για την ερμηνεία και την εξήγηση του φαινομένου συνάδουν με την νεοκλασική οικονομική προσέγγιση για την ορθολογική μεγιστοποίηση της χρησιμότητας σε υλικούς όρους υπό τις βασικές υποθέσεις της ορθολογικότητας και του ατομικού συμφέροντος. Οι θεωρίες αυτές περιλαμβάνουν την επιλογή βάσει γενετικής συγγένειας και την άμεση και έμμεση αμοιβαιότητα, που τονίζουν το ρόλο των ανταποδοτικών στρατηγικών και της δημιουργίας φήμης για την τήρηση συνεργατικών προτύπων. Ωστόσο οι θεωρίες αυτές δεν προσφέρουν επαρκή βάση εξήγησης για το φαινόμενο της συνεργασίας σε μη επαναλαμβανόμενες αλληλεπιδράσεις πολλών αγνώστων μεταξύ τους ατόμων σε συνθήκες ανωνυμίας όπου η πληροφόρηση είναι ελλιπής και η δημιουργία φήμης αδύνατη. Το κενό αυτό έρχεται να καλύψει η θεωρία της ισχυρής αμοιβαιότητας που δίνει ιδιάιτερο βάρος στο ρόλο που παίζουν οι κοινωνικές προτιμήσεις και οι κοινωνικοί και ηθικοί κανόνες στις αλληλεπιδράσεις των ατόμων. Κεντρικό στοιχείο της είναι η αλτρουιστική ή δαπανηρή τιμωρία, που αναφέρεται στην προθυμία των ατόμων να επωμίζονται υλικά κόστη προκειμένου να τιμωρήσουν μια μη συνεργατική συμπεριφορά παρ’ότι δεν έχουν να περιμένουν κάποιο άμεσο ή έμμεσο ίδιον υλικό όφελος από την πράξη αυτή. Η αλτρουιστική τιμωρία εντάσσεται στο πλαίσιο του βιολογικού ή συμπεριφορικού ορισμού του αλτρουισμού και οι καταβολές της σχετίζονται με την ύπαρξη κοινωνικών προτιμήσεων στους ανθρώπους. Ένας μεγάλος όγκος συμπεριφορικών πειραμάτων φωτίζει την παρουσία της ως αποτέλεσμα των ψυχολογικών μηχανισμών κινητοποίησης των ατόμων και αναδεικνύει τη σημασία της για την εδραίωση και τη σταθεροποίηση της συνεργασίας. Τα ευρήματα αυτά υποστηρίζονται και από νευροαπεικονιστικές μελέτες που δείχνουν ότι η αλτρουιστική τιμωρία προκαλεί συναισθήματα θετικής ανταμοιβής στα άτομα. Τέλος, ενδείξεις από εξελικτικά ανθρωπολογικά μοντέλα προσφέρουν επίσης περαιτέρω υποστήριξη για το ρόλο της στην εξελικτική σταθεροποίηση της συνεργασίας. Το σύνολο αυτών των δεδομένων καλεί για μια αναθεώρηση της νεοκλασικής θεωρίας χρησιμότητας, η οποία χρειάζεται να επεκταθεί προκειμένου να περιλαμβάνει και τις παραμέτρους των κοινωνικών προτιμήσεων των ατόμων. Ο ρόλος της αλτρουιστικής τιμωρίας στην ανάπτυξη συνεργατικών προτύπων πρέπει να αναγνωριστεί, ωστόσο ο βαθμός στον οποίο αποτελεί παράγοντα-κλειδί για την ανθρώπινη συνεργασία στις σύγχρονες κοινωνίες μένει ακόμη να εκτιμηθεί. Τα σύγχρονα συνεργατικά πρότυπα και οι κανόνες κοινωνικής συμπεριφοράς μοιάζει να έχουν εξελιχθεί και να να έχουν προοδευτικά βασιστεί σε λιγότερο δαπανηρούς μηχανισμούς επιβολής

κυρώσεων και αστυνόμευσης, όπως οι οργανωμένοι κοινωνικοί συνασπισμοί και το οργανωμένο κράτος δικαίου.

Λέξεις-κλειδιά

αλτρουιστική (δαπανηρή) τιμωρία, συνεργασία, αμοιβαιότητα, κοινωνικές προτιμήσεις, κανόνες κοινωνικής συμπεριφοράς

ABSTRACT

The puzzle of cooperation in human societies constitutes a central question in the social sciences, since its extent renders it a unique phenomenon in the animal kingdom. A large part of the theories developed in order to explain and interpret it are consistent with the neoclassical economic approach of utility maximization in material terms under the standard assumptions of self-interest and rationality. These theories include choice based on kin selection as well as direct and indirect reciprocity, which emphasize the role of reciprocal strategies and reputation building for the observation of cooperative norms. However, these theories do not offer a sufficient explanatory basis for the phenomenon of cooperation in one-shot interactions of a number of strangers under anonymity conditions where information is incomplete and reputation building is impossible. The theory of strong reciprocity attempts to cover this gap by emphasizing the role of social preferences and introducing the concept of altruistic or costly punishment. Altruistic punishers are willing to incur material costs in order to punish a non-cooperative behavior, even though they expect no direct or indirect own material benefit from this action. Altruistic punishment can be understood within the context of the biological or behavioral definition of altruism and its origins can be traced to the existence of social preferences in humans. A wide range of behavioral experiments elucidates its emergence as a result of the psychological motivation mechanisms of individuals and reveals its importance for the establishment and stabilization of cooperation. These findings are supported by neuroimaging studies that show that altruistic punishment elicits feelings of positive reward. Finally, evidence from evolutionary anthropological models also offers further support for its role in the evolutionary stabilization of cooperation. The above findings call for a review of the neoclassical utility theory approach, which needs to be extended in order to include parameters of human social and other-regarding preferences. The role of altruistic punishment in the development of cooperative norms must be acknowledged, however the degree to which it constitutes a key factor for human cooperation in modern societies has yet to be assessed. Modern cooperative and social norms seem to have evolved and become progressively based on less costly mechanisms of sanctioning and policing, like coordinated social coalitions, common pool institutions and the rule of law.

Keywords

altruistic (costly) punishment, cooperation, reciprocity, social preferences, social norms

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΛΗΨΗ 3
Λέξεις-κλειδιά 4
ABSTRACT 5
Keywords 5
ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ 8
ΑΠΟΔΟΣΗ ΟΡΩΝ 9
ΕΙΣΑΓΩΓΗ 11
Σκοπός, δομή και μέθοδος 15
Κεφάλαιο 1: Θεωρίες για την ανάπτυξη της συνεργασίας και του αλτρουισμού 17
1.1 Επιλογή βάσει γενετικής συγγένειας 18
1.2 Άμεση αμοιβαιότητα 19
1.3 Έμμεση αμοιβαιότητα 20
1.4 Ισχυρή αμοιβαιότητα και αλτρουιστική τιμωρία 21
Κεφάλαιο 2: Η αλτρουιστική τιμωρία: ενδείξεις και ευρήματα από πειράματα και μελέτες 25
2.1 Συμπεριφορικά πειράματα 26
2.1.1 Ultimatum game (one-shot) 27
2.1.2 Ultimatum (με δημιουργία φήμης) 29
2.1.3 Τιμωρία από τρίτο πρόσωπο 32
2.1.4. Παίγνια δημόσιων αγαθών 36
2.2 Κοινωνικές προτιμήσεις στους ανθρώπους 45
2.2.1 Κοινωνικές προτιμήσεις 45
2.2.2 Πειραματικά ευρήματα 46
2.3 Νευροβιολογικές ενδείξεις 47
2.4 Εξελικτικά μοντέλα 52
2.4.1 Πολιτισμική-γονιδιακή συνεξέλιξη 52
2.4.2 Πολιτισμική ομαδική επιλογή 56
Κεφάλαιο 3: Αλτρουιστική τιμωρία: εξωτερική εγκυρότητα 58
3.1 Απόκλιση πειραματικών συνθηκών από τις πραγματικές 59
3.1.1 Δυνατότητα επικοινωνίας 60
3.1.2 Μορφή τιμωρίας 61
3.1.3 Δυνατότητα αποχώρησης από το παιχνίδι 61

3.1.4 Δυνατότητα αντι-τιμωρίας 62
3.1.5 Κόστος τιμωρίας και συνολικές απολαβές 63
3.2 Περιορισμοί της εφαρμογής και του ρόλου της αλτρουιστικής τιμωρίας 63
3.2.1 Περιορισμοί στο εγγύς και απώτατο επίπεδο 63
3.2.2 Θεσμοί κοινών πόρων (common pool resources) 64
3.2.3 Αλτρουιστική ανταμοιβή 65
ΣΥΜΠΕΡΑΣΜΑΤΑ 67
ΒΙΒΛΙΟΓΡΑΦΙΑ 71

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

Σχήμα 1α, 1β: Κατώφλια αποδοχής των δεκτών στο ultimatum game με και χωρίς δυνατότητα δημιουργίας φήμης. 31
Σχήμα 2: Αλτρουιστική τιμωρία απο τρίτα μέρη που δεν επηρεάζονται άμεσα από την παραβίαση ενός κοινωνικού κανόνα δίκαιης κατανομής.. 35
Σχήμα 3: Πόντοι ποινής που δέχτηκαν τα άτομα σε συνάρτηση με την απόκλιση της συνεισφοράς τους από τη μέση συνεισφορά των άλλων μελών της ομάδας. 42
Σχήμα 4: Μέση συνεισφορά κατά την εξέλιξη των περιόδων σε παίγνιο δημόσιου αγαθού με σταθερή σύνθεση ομάδας (10 ομάδες). 43
Σχήμα 5: Μέση συνεισφορά κατά την εξέλιξη των περιόδων σε παίγνιο δημόσιου αγαθού με τυχαία σύνθεση ομάδας (18 ομάδες). 44

ΑΠΟΔΟΣΗ ΟΡΩΝ

στιγμιαία αλληλεπίδραση	one-shot interaction
αλτρουιστική ανταμοιβή	altruistic rewarding
αλτρουιστική τιμωρία	altruistic punishment
άμεση αμοιβαιότητα	direct reciprocity
αμοιβαίος αλτρουισμός	reciprocal altruism
αντι-τιμωρία	counter-punishment
απλά συνεργατικό άτομο	contributor
απώτερο επίπεδο	ultimate level
αρμοστικότητα	fitness
ατελής σύμβαση	incomplete contract
δαπανηρή τιμωρία	costly punishment
δεσμευτική σύμβαση	binding contract
εγγύς επίπεδο	proximate level
εγωιστική συμπεριφορά	self-regarding behavior
έμμεση αμοιβαιότητα	indirect reciprocity
ενσυναίσθηση	empathy
εξωτερική εγκυρότητα	external validity
επιβίωση του καταλληλότερου	survival of the fittest
επιλογή βάσει γενετικής συγγένειας	genetical kinship theory, kin selection
εσωτερικοποίηση κοινωνικών κανόνων	internalization of norms
ευρετικές μέθοδοι	heuristics
θεσμοί κοινών πόρων	common pool institutions
θεωρία χρησιμότητας	utility theory
ισχυρή αμοιβαιότητα	strong reciprocity
κάθετη μετάδοση	vertical transmission
κανόνες κοινωνικής συμπεριφοράς	social norms
κοινωνικές προτιμήσεις	social preferences
κοινωνική μάθηση	social learning
κοινωνικοποίηση	socialization

λαθρεπιβάτης	free rider
Λειτουργική Μαγνητική Τομογραφία	Functional Magnetic Resonance Imaging, fMRI
μετάδοση βάσει απολαβών ή κύρους	payoff or prestige based transmission
μετάδοση βάσει συμμόρφωσης στα καθιερωμένα	conformist based transmission
νεοκλασικά οικονομικά	neoclasical economics
ομαδική επιλογή σε πολλαπλά επίπεδα	multilevel group selection
οπισθοβατική επαγωγή	backward induction
ορθολογικοί μεγιστοποιητές	rational utility maximizers
οριζόντια μετάδοση	horizontal transmission
παίγνιο δημόσιου αγαθού	public good game
παίγνιο δικτάτορα	dictator game
παίγνιο μιας ευκαιρίας	one-shot game
παίγνιο τελεσίγραφου	ultimatum game
παραβάτης συνεργατικού προτύπου	defector
πολιτισμική ομαδική επιλογή	cultural group selection
πολιτισμική-γονιδιακή συνεξέλιξη	gene-culture coevolution
προδοσία συνεργασίας	defection
προτιμήσεις που αφορούν τους άλλους	other-regarding preferences
κοστοβόρος σηματοδότηση	costly signaling
συμπεριφορική θεωρία παιγνίων	behavioral game theory
τιμωρία από δεύτερα μέρη	second-party punishment
τιμωρία από τρίτα μέρη	third-party punishment
τιμωρία δεύτερης τάξης	second order punishment
τιμωρός	punisher
Τομογραφίας Εκπομπής Ποζιτρονίων	Positron Emission Tomography, PET
φήμη	reputation
φυσική επιλογή εντός μιας ομάδας	within group selection
φυσική επιλογή μεταξύ των ομάδων	between group selection
χρησιμότητα	utility

ΕΙΣΑΓΩΓΗ

Το πρόβλημα της συνεργασίας αποτελεί ένα θεμελιώδες ερώτημα στις κοινωνικές και βιολογικές επιστήμες. Η εκτεταμένη συνεργασία σε πολλαπλά επίπεδα αποτελεί τη βάση του πολιτισμού και η εκδήλωσή της προσδιορίζει την ίδια τη φύση του ανθρώπινου είδους. Η εξήγηση των μηχανισμών που τη στηρίζουν και η ανάπτυξη θεωριών για το πώς αυτή εξελίχθηκε έχει λοιπόν μεγάλη σημασία. Η βιβλιογραφική αυτή μελέτη πραγματεύεται το ρόλο που μπορεί να παίξει στην εδραίωση και την εξέλιξη της συνεργασίας και των κοινωνικών κανόνων η αλτρουιστική τιμωρία, μια συμπεριφορά κατά την οποία τα άτομα τιμωρούν παραβιάσεις των συνεργατικών προτύπων με κόστος για τα ίδια, χωρίς να περιμένουν κάποιο άμεσο ή έμμεσο υλικό όφελος.

Όταν μιλάμε για συνεργασία εννοούμε την εκδήλωση συμπεριφορών από ένα άτομο που επιφέρουν οφέλη σε άλλα άτομα (West et al., 2007). Όταν τα άτομα σε μια ομάδα συνεργάζονται μπορούν να πετύχουν οφέλη που υπερβαίνουν το άθροισμα των μεμονωμένων συμβολών των μερών. Η έννοια και η εκδήλωση της συνεργασίας συνδέεται στενά με τους κανόνες κοινωνικής συμπεριφοράς.

Οι κανόνες κοινωνικής συμπεριφοράς (*social norms*), ή απλούστερα κοινωνικοί κανόνες, είναι κανονιστικά πρότυπα συμπεριφοράς που επιβάλλονται μέσω άτυπων κοινωνικών κυρώσεων. Πρόκειται για ένα σύνολο κοινών προτύπων πεποιθήσεων, ιδεών, πρακτικών και συμπεριφορών που αποκτώνται μέσω της κοινωνικής μάθησης (*social learning*) (Boyd & Richerson, 2005). Οι κανόνες αυτοί καθορίζουν την αναμενόμενη και κοινωνικά αποδεκτή συμπεριφορά για ένα πλήθος καταστάσεων και σχέσεων μεταξύ των μελών μιας κοινωνίας. Μπορεί να αφορούν υποχρεώσεις για συνεργατικές συμπεριφορές στα πλαίσια μιας ομάδας, συμμετοχή σε κοινές προσπάθειες, αντιλήψεις περί δίκαιης κατανομής αγαθών, σεβασμό στα άλλα μέλη της κοινωνίας. Η απειλή της επιβολής κυρώσεων όταν οι κανόνες αυτοί παραβιάζονται εγγυάται την ενίσχυση και την εδραίωσή τους. Η διερεύνηση επομένως των μηχανισμών που οδηγούν στην επιβολή αυτών των κυρώσεων είναι κρίσιμη προκειμένου να εντοπίσουμε τις βάσεις της συνεργασίας στις ανθρώπινες κοινωνίες.

Συνεργασία στις ανθρώπινες κοινωνίες

Οι ανθρώπινες κοινωνίες διαθέτουν μια μοναδικότητα σε σχέση με όλα τα υπόλοιπα είδη του ζωικού βασιλείου. Οι άνθρωποι που τις αποτελούν συνεργάζονται μεταξύ τους συστηματικά και

σε μεγάλες κλίμακες, συχνά προσφέρουν βοήθεια σε ξένους και πολλές φορές συμβάλλουν στη βελτίωση της ευημερίας άλλων ατόμων ακόμα και με κόστος για τους ίδιους (Nowak & Highfield, 2011). Η συμπεριφορά τους ακολουθεί πρότυπα συμμόρφωσης με μια σειρά κοινωνικών κανόνων. Οι ανθρώπινες κοινωνίες είναι βασισμένες σε ένα λεπτομερή καταμερισμό της εργασίας μεταξύ γενετικά μη συσχετισμένων ατόμων σε μεγάλες ομάδες. Αυτό είναι προφανές στις σύγχρονες κοινωνίες με τους μεγάλους οργανισμούς και τα οργανωμένα κράτη, ισχύει ωστόσο και σε πιο πρωτόγονες κοινωνίες κυνηγών, τροφοσυλλεκτών και νομάδων κτηνοτρόφων. Σε αυτές τις κοινωνίες παρατηρούνται δίκτυα ανταλλακτικών σχέσεων και οργανωμένες πρακτικές για το μοίρασμα της τροφής, το συνεργατικό κυνήγι και τη διεξαγωγή πολέμου (Henrich & Henrich, 2007).

Η συνεργασία βέβαια παρατηρείται και σε άλλα είδη ζώων, όπου όμως περιορίζεται σε μικρές ομάδες γενετικά σχετισμένων ατόμων (Silk, 2009). Ακόμα και σε κοινωνίες πρωτευόντων, που είναι οι πιο κοντινοί εξελικτικά συγγενείς με το ανθρώπινο είδος, η συνεργασία είναι κατά πολλές τάξεις μεγέθους λιγότερο ανεπτυγμένη. Τα άλλα είδη στα οποία παρατηρούνται εκτεταμένες συνεργατικές δομές είναι τα κοινωνικά έντομα, όπως τα μυρμήγκια και οι μέλισσες, όμως και σε αυτά καθοριστικός παράγοντας είναι η γενετική συγγένεια. Γενικά η συνεργασία στα ζώα περιορίζεται σε γενετικά συσχετιζόμενα άτομα και μέχρι ένα σημείο σε ανταποδοτικές πράξεις αμοιβαιότητας, στη βάση του *quid pro quo*: το ένα ζώο προσφέρει βοήθεια π.χ. στη σωματική φροντίδα ή την εύρεση τροφής σε ένα άλλο, όταν και το άλλο ανταποδίδει αυτή ή κάποια άλλη βοηθητική πράξη. Ένα τέτοιο παράδειγμα είναι η ανταλλαγή τροφής που έχει παρατηρηθεί σε ένα είδος νυχτερίδων. Νυχτερίδες που έχουν βρει τροφή μεταφέρουν ένα τμήμα από το γεύμα τους σε μέλη της ομάδας που δεν τα κατάφεραν, όταν τα μέλη αυτά τις έχουν βοηθήσει με τον ίδιο τρόπο στο παρελθόν (Wilkinson, 1984). Τέτοιου τύπου περιορισμένες ανταλλαγές σε ανταποδοτική βάση παρατηρούνται και σε άλλα είδη, ωστόσο απέχουν πολύ από τα πολύπλοκα και υψηλά επίπεδα συνεργασίας των ανθρώπινων κοινωνιών.

Θεωρητικό πλαίσιο για τη συνεργασία

Το φαινόμενο της συνεργασίας μπορεί να προσεγγιστεί μέσα από θεωρίες που αναπτύσσονται στα πλαίσια των κοινωνικών επιστημών, της ψυχολογίας, της βιολογίας, της ανθρωπολογίας και των οικονομικών. Η μεγάλη πρόκληση έγκειται στην εξήγηση του πώς μια συμπεριφορά που έχει κόστος για το άτομο και αυξάνει το όφελος άλλων ατόμων μπορεί να διατηρηθεί από τη

φυσική επιλογή και να έχει εξελικτική σταθερότητα. Η ανεύρεση των γνωσιακών διαδικασιών και μηχανισμών που βρίσκονται πίσω από τη συμπεριφορά αυτή, καθώς και η μελέτη σχετικά με την εξελικτική τους προέλευση έχει ιδιαίτερο ενδιαφέρον για τη γνωσιακή επιστήμη.

Η συμπεριφορά και οι επιλογές των ανθρώπων σε συνθήκες αλληλεπίδρασης με άλλα άτομα όπου υπάρχει αναμονή απολαβών έχει μελετηθεί εκτενώς στα πλαίσια της οικονομικής επιστήμης. Σε όλη τη διάρκεια της ιστορίας της κλασικής οικονομικής σκέψης το κύριο ρεύμα των θεωριών που αναπτύχθηκαν στηρίχθηκε σε μια κεντρική υπόθεση: οι άνθρωποι είναι ορθολογικοί μεγιστοποιητές (*rational utility maximizers*). Σύμφωνα με τη νεοκλασική θεωρία, το κίνητρο για τις επιλογές των ατόμων είναι η μεγιστοποίηση των υλικών τους απολαβών. Παίρνουν αποφάσεις ορθολογικά με γνώμονα το ατομικό τους συμφέρον και επιδιώκουν πάντα να μεγιστοποιήσουν τις απολαβές τους (με την ευρεία έννοια) επωμιζόμενοι το μικρότερο δυνατό κόστος, κάτι που αναφέρεται στη βιβλιογραφία και ως ορθολογικός εγωισμός (Glimcher et al., 2009). Η θεώρηση αυτή των ανθρώπων συνδέθηκε με τη δαρβινική θεωρία της εξέλιξης και την "πάλη για την ύπαρξη", όπως αποκαλούσε ο ίδιος ο Δαρβίνος τη διαδικασία όπου μέσω της φυσικής επιλογής ευνοείται η μεταβίβαση των καταλληλότερων χαρακτηριστικών των οργανισμών από γενιά σε γενιά. Η "επιβίωση του καταλληλότερου" (*survival of the fittest*) φέρεται να υποδεικνύει πως σε ό,τι αφορά τους ανθρώπους η εξέλιξη ευνοεί τα άτομα εκείνα που υιοθετούν συμπεριφορές που προωθούν το στενό υλικό ατομικό τους συμφέρον.

Η θεώρηση αυτή αποτυγχάνει ωστόσο να εξηγήσει ένα από τα βασικότερα χαρακτηριστικά των ανθρώπινων κοινωνιών: την ευρέως διαδεδομένη συνεργασία μεταξύ των ατόμων και την ύπαρξη αλτρουιστικών συμπεριφορών (Henrich & Henrich, 2007). Νέες οικονομικές θεωρήσεις που επιχειρούν να ενσωματώσουν στο μοντέλο των ανθρώπων προτιμήσεων και άλλους παράγοντες πέρα από τους αυστηρά οικονομικούς, όπως οι θεωρίες των κοινωνικών προτιμήσεων (*social preference theories*) (Falk & Fischbacher, 2006; Charness & Rabin, 2000) μπορούν να προσφέρουν ένα καλύτερο υπόβαθρο για την εξήγηση των συνεργατικών συμπεριφορών. Η εκτενής βιβλιογραφία και έρευνα στην οικονομική επιστήμη γύρω από τις προτιμήσεις και οι συμπεριφορικές προσεγγίσεις σχετικά με αυτές παρέχουν έναν πλούσιο όγκο πληροφοριών και ενδείξεων σχετικά με τους γνωσιακούς μηχανισμούς που υποκινούν συμπεριφορές που υποστηρίζουν τη συνεργασία. Για το λόγο αυτό οι ενδείξεις σχετικά με τις εγωιστικές ή αλτρουιστικές από οικονομικής άποψης προτιμήσεις και επιλογές των ατόμων θα αποτελέσουν βασικό κορμό για αυτή την εργασία.

Η αναφορά στον αλτρουισμό και τις αλτρουιστικές πράξεις σε αυτή τη βιβλιογραφική μελέτη ακολουθεί τον ορισμό του αλτρουισμού κατά τη βιολογική ή συμπεριφορική έννοια. Ως συμπεριφορικό αλτρουισμό ορίζουμε τις πράξεις ενός ατόμου που έχουν ένα κόστος για το ίδιο και επιφέρουν οικονομικά οφέλη σε άλλα άτομα (Batson, 1991). Ο ορισμός αυτός, σε αντίθεση με τον ψυχολογικό, δεν εξετάζει αν τα κίνητρα πίσω από την πράξη είναι επίσης αλτρουιστικά και δεν απαιτεί να μην υπάρχει κάποια ηδονική ή συναισθηματική ανταμοιβή για το άτομο από την αλτρουιστική πράξη. Αυτό που εξετάζεται είναι το αποτέλεσμα της πράξης σε υλικούς όρους για το άτομο και τους άλλους και όχι αν τα άτομα λαμβάνουν κάποιου είδους εσωτερική ανταμοιβή σε ψυχολογικούς-συναισθηματικούς όρους.

Η μελέτη του φαινομένου της συνεργασίας και των γνωσιακών διαδικασιών που οδηγούν τα άτομα σε συμπεριφορές που τη στηρίζουν μπορεί να υποβοηθηθεί αν χωρίσουμε την ανάλυσή της σε δύο επίπεδα: το εγγύς (*proximate level*) και το απώτερο (*ultimate level*) επίπεδο (Henrich & Henrich, 2007). Στο εγγύς επίπεδο αναλύονται οι ψυχολογικοί μηχανισμοί, οι προτιμήσεις, οι πεποιθήσεις, οι αξίες και τα κίνητρα που ωθούν την ανθρώπινη συμπεριφορά, τη λήψη αποφάσεων και τις πράξεις των ατόμων. Στο απώτερο επίπεδο αναλύονται οι εξελικτικές διαδικασίες που οδήγησαν στη διαμόρφωση των ανωτέρω.

Έτσι λοιπόν μπορούμε να έχουμε μια διπλή προσέγγιση στο φαινόμενο της συνεργασίας από δύο σκοπιές :τη συμπεριφορική και την εξελικτική. Η πρώτη προσέγγιση (εγγύς επίπεδο) μπορεί να μας δώσει πληροφορία για τον τρόπο με τον οποίο λειτουργεί η συνεργασία στις σύγχρονες ανθρώπινες κοινωνίες και ομάδες ατόμων και για τα κίνητρα και τους ψυχολογικούς και γνωσιακούς μηχανισμούς που στηρίζουν την εκδήλωση συνεργατικών συμπεριφορών. Η επιστήμη της ψυχολογίας, η κοινωνιολογία, τα οικονομικά μπορούν να συμβάλουν στην προσέγγιση αυτή ρίχνοντας φως στους μηχανισμούς και τις δυναμικές που αναπτύσσονται στις αλληλεπιδράσεις μεταξύ των ατόμων.

Η δεύτερη προσέγγιση (απώτερο επίπεδο) είναι επίσης απολύτως απαραίτητη προκειμένου να κατανοήσουμε πλήρως το μοναδικό αυτό φαινόμενο. Για να καταλάβουμε τη σημασία και τη λειτουργία της συνεργασίας και των υποστηρικτικών της μηχανισμών στις σύγχρονες κοινωνίες πρέπει να μελετήσουμε το πώς αυτή αναδύθηκε και αναπτύχθηκε στην πορεία της ανθρώπινης εξέλιξης. Η εξελικτική ανθρωπολογία και βιολογία μπορούν να μας προσφέρουν τα εργαλεία που χρειαζόμαστε για το σκοπό αυτό. Είναι πολύ σημαντικό να μπορούμε να διαμορφώσουμε μια θεωρία σχετικά με τις ρίζες και τους λόγους εδραίωσης της συνεργασίας, ώστε να μπορούμε

να αναλύσουμε καλύτερα το ρόλο που έπαιξε στην εξέλιξη του ανθρώπινου είδους και τον τρόπο με τον οποίο αυτός αποτυπώθηκε στη συγκρότηση των σχετικών γνωσιακών διαδικασιών.

Ένα θεωρητικό πλαίσιο λοιπόν για την εξήγηση του φαινομένου της συνεργασίας και της συμμόρφωσης με κοινωνικούς κανόνες πρέπει να διαθέτει ερείσματα και αναφορές και στα δύο αυτά επίπεδα. Οι παραδοσιακές θεωρητικές προσεγγίσεις στο ζήτημα περιλαμβάνουν θεωρίες που βασίζονται σε άμεσες και έμμεσες μορφές αμοιβαιότητας, σύμφωνα με τις οποίες η συνεργασία βασίζεται στα άμεσα ή έμμεσα υλικά οφέλη που προσφέρει στα άτομα, κάτι που είναι σύμφωνο με τη θεωρία μεγιστοποίησης του εγωιστικού υλικού συμφέροντος (Fehr & Fischbacher, 2003). Ωστόσο οι θεωρίες αυτές αποτυγχάνουν να εξηγήσουν συνεργατικές συμπεριφορές σε μια γκάμα συνθηκών, όπως οι μη επαναλαμβανόμενες αλληλεπιδράσεις με ξένους και οι αλληλεπιδράσεις σε συνθήκες ανωνυμίας. Έτσι μια ακόμη βάση για την εδραίωση συνεργατικών συμπεριφορών έχει προταθεί: πρόκειται για την αλτρουιστική τιμωρία (*altruistic punishment*), που αποτελεί συστατικό στοιχείο της θεωρίας της ισχυρής αμοιβαιότητας (Gintis, 2000; Fehr et al., 2002). Όταν αναφερόμαστε στην τιμωρία γενικά, εννοούμε την ανάληψη πράξεων ή επιλογών από ένα άτομο που οδηγούν σε μείωση των οικονομικών απολαβών άλλων ατόμων. Η αλτρουιστική τιμωρία ειδικά, απαιτεί την ανάληψη κόστους από τον τιμωρό και δεν του αποφέρει κάποιο άμεσο ή έμμεσο υλικό όφελος. Συμπεριφορικά πειράματα έχουν δείξει ότι η επιβολή της στα μη συνεργατικά άτομα έχει σαν αποτέλεσμα την ενίσχυση των συνεργατικών συμπεριφορών σε μια ομάδα (Camerer & Thaler, 1995; Fehr & Gächter, 2000). Προκύπτει λοιπόν το ερώτημα σχετικά με το ποιοι μπορεί να είναι οι μηχανισμοί που την υποκινούν και κατά πόσο η αλτρουιστική τιμωρία μπορεί να παίξει ρόλο στην ανάπτυξη και την εδραίωση της συνεργασίας.

Σκοπός, δομή και μέθοδος

Σκοπός αυτής της βιβλιογραφικής μελέτης είναι να παρουσιάσει τα στοιχεία γύρω από την ύπαρξη της αλτρουιστικής τιμωρίας και το ρόλο που αυτή μπορεί να έχει παίξει για την εξέλιξη και τη διατήρηση της συνεργασίας και των κοινωνικών κανόνων σε ένα πληθυσμό. Ο στόχος είναι να δοθούν ενδείξεις για την παρουσία της μεταξύ των συμπεριφορικών προτύπων των ατόμων, καθώς και για τους γνωσιακούς μηχανισμούς που μπορεί να την υποκινούν στο εγγύς

επίπεδο. Αναζητούνται επίσης ενδείξεις που μπορούν να δικαιολογήσουν εξελικτικά την ύπαρξή της στο απώτατο επίπεδο. Τέλος, ελέγχεται ο βαθμός και ο τρόπος με τον οποίο μπορεί να συμβάλει στην ανάπτυξη και την εδραίωση της συνεργασίας και των κοινωνικών κανόνων, καθώς και τα πιθανά όρια αυτής της επίδρασης.

Η μέθοδος που ακολουθήθηκε κατά τη συγγραφή της μελέτης ήταν η έρευνα σε βιβλία και άρθρα της σχετικής βιβλιογραφίας. Η συλλογή του υλικού έγινε με βάση τους σκοπούς και τους στόχους της έρευνας, καθώς και με αξιολόγηση των πρόσθετων πληροφοριών και συμπερασμάτων που προέκυψαν από την αρχική ανασκόπηση της βιβλιογραφίας.

Αρχικά θα παρουσιάσουμε τις παραδοσιακές θεωρίες για την εδραίωση της συνεργασίας και θα εκθέσουμε τους περιορισμούς τους και την απάντηση της θεωρίας της ισχυρής αμοιβαιότητας σε αυτούς. Στη συνέχεια θα παρουσιάσουμε ευρήματα από πειράματα και μελέτες που προσφέρουν ενδείξεις για την ύπαρξη της αλτρουιστικής τιμωρίας και την επίδρασή της στην ενίσχυση της συνεργασίας στο εγγύς επίπεδο. Οι ενδείξεις αυτές προέρχονται από συμπεριφορικά πειράματα και από νευροαπεικονιστικές έρευνες, οι οποίες υποδεικνύουν μια βιολογική βάση για την εκδήλωση της αλτρουιστικής τιμωρίας. Ο συνδυασμός των πειραματικών ευρημάτων θα μας οδηγήσει στην εξαγωγή συπερασμάτων σχετικά με τους γνωσιακούς μηχανισμούς που βρίσκονται πίσω από την εκδήλωση της συμπεριφοράς αυτής. Επίσης θα παρουσιάσουμε περαιτέρω υποστηρικτικές ενδείξεις για τον ρόλο της αλτρουιστικής τιμωρίας στην εξέλιξη της συνεργασίας στο απώτατο επίπεδο μέσα από εξελικτικά μοντέλα, καθώς και για τη σχέση που έχει με τις κοινωνικές προτιμήσεις. Ακολούθως θα επιχειρήσουμε μια παρουσίαση των πιθανών ορίων της αλτρουιστικής τιμωρίας και μια κριτική ματιά στο ρόλο που μπορεί να παίζει για τη συνεργασία στις συνθήκες του πραγματικού κόσμου. Τέλος, θα παραθέσουμε μια σύνοψη των ευρημάτων και των προβληματισμών που παρουσιάστηκαν και θα εκθέσουμε τα συμπεράσματα για τη φύση και το ρόλο της αλτρουιστικής τιμωρίας, καθώς και για τα όρια των συνθηκών εντός των οποίων μπορεί να αποτελέσει έναν ενισχυτικό παράγοντα για τη συνεργασία.

Κεφάλαιο 1: Θεωρίες για την ανάπτυξη της συνεργασίας και του αλτρουισμού

Επιστήμονες από διάφορους κλάδους έχουν προσπαθήσει να δώσουν απάντηση στο ερώτημα του πώς εξελίχθηκε η συνεργασία και οι κανόνες κοινωνικής συμπεριφοράς και του ποιοι είναι οι βασικοί μηχανισμοί που τους σταθεροποιούν και τους στηρίζουν στις ανθρώπινες κοινωνίες. Στις καθημερινές συναλλαγές μεταξύ των ατόμων, ακόμα και στις σύγχρονες κοινωνίες με τις οργανωμένες δομές και θεσμούς για τη στήριξη της συνεργασίας όπως οι νόμοι, τα δικαστήρια και η αστυνομία, υπάρχει πάντα ένα υλικό κίνητρο για παραβίαση των κανόνων και εκμετάλλευση των άλλων. Η συμπεριφορά του λαθρεπιβάτη (*free-riding*), όπου το άτομο επιλέγει να εξαπατήσει τα άλλα μέρη της συναλλαγής, δρέποντας τους καρπούς των κόπων τους χωρίς το ίδιο να συνεισφέρει, είναι συμφέρουσα από μια στενά εγωιστική πλευρά. Η αθέτηση των υποχρεώσεων μιας συμφωνίας μπορεί να αυξήσει τις υλικό όφελος του ατόμου, ιδιαίτερα αν οι συνθήκες της συναλλαγής είναι τέτοιες που το άτομο δε μπορεί να εξαναγκαστεί με κάποιο τρόπο στην τήρηση των δεσμεύσεών του προκειμένου να δρέψει τις δικές του απολαβές από τη συμφωνία. Οι συμφωνίες αυτές της μορφής ονομάζονται ατελείς συμβάσεις (*incomplete contracts*) και αποτελούν το μεγαλύτερο μέρος των συμφωνιών συναλλαγής μέσα στις κοινωνίες (Hart & Moore, 1998). Ακόμα και όταν οι συμφωνίες σε μεγάλο βαθμό είναι νομικά δεσμευτικές (*binding contracts*), σχεδόν πάντα υπάρχουν κάποια στοιχεία τους όπου ο έλεγχος είναι ατελής και η τήρηση των συμπεφωνημένων βασίζεται στην καλή πρόθεση των συναλλασσόμενων μερών. Στις αρχέγονες κοινωνίες αυτό θα ίσχυε για την πλειοψηφία των συναλλαγών, καθώς απουσίαζαν οι οργανωμένες δομές ελέγχου που υπάρχουν σήμερα. Σύμφωνα λοιπόν με την κλασική θεωρία μεγιστοποίησης της χρησιμότητας, τα άτομα θα έπρεπε να επιλέγουν πάντα να αθετήσουν τις συνεργατικές τους υποχρεώσεις όταν αυτό είναι δυνατό, καθώς έτσι αυξάνουν τις ατομικές τους απολαβές.

Στην πραγματικότητα όμως αυτό που παρατηρούμε στις καθημερινές συναλλαγές των ατόμων είναι η τήρηση σε μεγάλο βαθμό των υποχρεώσεων των ατελών συμβάσεων. Τα άτομα συχνά επιλέγουν να τηρήσουν τις υποχρεώσεις τους σε μια συνεργατική συμφωνία και περιμένουν από τα άλλα μέρη να κάνουν το ίδιο. Όταν υπάρχει αθέτηση από κάποιο συναλλασσόμενο μέρος, τα άλλα άτομα συχνά αποζητούν την επιβολή κάποιας ποινής που στόχο έχει την αποστέρηση των ‘παράνομων’ οφελών που αποκόμισε ο παραβάτης και την

επανόρθωση του συνεργατικού προτύπου. Η συμπεριφορά αυτή έρχεται σε αντίθεση με την υπόθεση των απολύτως εγωιστικών ατόμων που ενδιαφέρονται μόνο για το ατομικό τους συμφέρον και τις ατομικές τους απολαβές, καθώς πολλές φορές τα άτομα πρέπει να επωμιστούν ένα υλικό κόστος προκειμένου να επιβάλουν την τιμωρία.

Η εξήγηση αυτής της συμπεριφοράς θέτει προκλήσεις για το κλασικό οικονομικό μοντέλο και έχει ενδιαφέρον από γνωσιακής πλευράς. Αφενός η πρόκληση αφορά το εγγύς επίπεδο δικαιολόγησης των κινήτρων που βρίσκονται πίσω από αυτήν παρατηρούμενη συμπεριφορά των ατόμων σήμερα. Αφετέρου αφορά και το απώτερο επίπεδο δικαιολόγησης των εξελικτικών μηχανισμών που επέτρεψαν και ευνόησαν την επιβίωσή των χαρακτηριστικών που οδηγούν τα άτομα στην υιοθέτησή της. Στην προσπάθεια να ερμηνευθεί η συμπεριφορά αυτή εντός του πλαισίου της ορθολογικής μεγιστοποίησης της χρησιμότητας διάφορες θεωρίες αναπτύχθηκαν.

1.1 Επιλογή βάσει γενετικής συγγένειας

Η παλαιότερη θεωρία είναι αυτή της *επιλογής βάσει γενετικής συγγένειας (genetical kinship theory, kin selection)* (Hamilton, 1964). Σύμφωνα με αυτή, τα άτομα επωμίζονται κόστη προκειμένου να αυξήσουν τα οφέλη ατόμων με τα οποία μοιράζονται γενετική συγγένεια. Συνήθως, και ιδιαίτερα στο αρχέγονο περιβάλλον, τα άτομα που σχετίζονται γενετικά σχηματίζουν ομάδες στενής συνεργασίας, όπου είναι συχνές οι πράξεις που αποφέρουν οφέλη στα άλλα μέλη της ομάδας, ακόμα και όταν οι ίδιες φέρουν κόστος για το άτομο που ενεργεί.

Η συμπεριφορά αυτή είναι συμβατή με τη θεωρία της εγωιστικής υλικής μεγιστοποίησης, εφόσον τα άτομα έχουν υλικό συμφέρον να βοηθήσουν τους συγγενείς τους, καθώς έχουν κάθε λόγο να περιμένουν ανταπόδοση των πράξεών τους. Έχουν δηλαδή να περιμένουν κάποιο υλικό όφελος στο βραχυπρόθεσμο και μακροπρόθεσμο μέλλον από την πράξη τους, είναι επομένως λογικό να δέχονται ένα βραχυπρόθεσμο υλικό κόστος προκειμένου να αποκομίσουν μεγαλύτερα μελλοντικά οφέλη. Το υλικό συμφέρον όλων είναι να τηρήσουν τα συνεργατικά πρότυπα και να τιμήσουν την εμπιστοσύνη των υπόλοιπων μελών της ομάδας, στηρίζοντας το πνεύμα της συνεργασίας. Τα άτομα γνωρίζονται καλά μεταξύ τους και οι δεσμοί που αναπτύσσουν μεταξύ συγγενών διαρκούν τις περισσότερες φορές για το σύνολο του βίου τους. Ο μακρύς ορίζοντας των σχέσεων και ο υψηλός βαθμός πληροφόρησης σε σχέση με την ταυτότητα και το ιστορικό συμπεριφοράς των ατόμων καθιστά τη συνεργασία την πιο επικερδή μορφή αλληλεπίδρασης

(West et al., 2002; Sachs et al., 2004).

Στο πλαίσιο αυτό και η επιβολή τιμωρίας, ακόμα και με κόστος για τον τιμωρό, είναι συμβατή με την εγωιστική μεγιστοποίηση, εφόσον λειτουργεί προστατευτικά για τη συνεργασία που αυξάνει τις ατομικές απολαβές. Από εξελικτικής σκοπιάς επίσης, η συμπεριφορά αυτή μπορεί να εξηγηθεί, εφόσον πέρα από τη βελτίωση της αρμοστικότητας των ατόμων, είναι ευνοϊκή και για την αρμοστικότητα των συγγενών. Οι συγγενείς μοιράζονται γενετικό υλικό και επομένως τα τα άτομα που φέρουν χαρακτηριστικά που τους δημιουργούν συνεργατικές τάσεις δέχονται μια πολλαπλή ευνοϊκή εξελικτική πίεση.

Έτσι, στο πλαίσιο ομάδων που αποτελούνται από συγγενικά άτομα η συνεργασία μπορεί να εξηγηθεί με τους όρους της εγωιστικής υλικής μεγιστοποίησης και είναι βέβαιο ότι αυτή η κινητήριος δύναμη έπαιξε πολύ σημαντικό ρόλο στην εξέλιξή της. Συνεργασία στη βάση της γενετικής συγγένειας έχει παρατηρηθεί σε πολλά είδη ζώων, με αποκορύφωμα τις οργανωμένες δομές των κοινωνικών εντόμων (Silk, 2009). Το γεγονός αυτό αποτελεί ένδειξη ότι εξελικτικά η επιλογή βάσει γενετικής συγγένειας αποτέλεσε ίσως το πρώτο υπόβαθρο για την ανάπτυξη και την εδραίωση της συνεργασίας.

1.2 Άμεση αμοιβαιότητα

Στην καθημερινή ζωή στις ανθρώπινες κοινωνίες όμως, η συνεργασία δεν περιορίζεται σε γενετικά σχετιζόμενα μεταξύ τους άτομα. Οι άνθρωποι πραγματοποιούν καθημερινά ένα πλήθος συναλλαγών με άλλους ανθρώπους με τους οποίους δε μοιράζονται γενετικά χαρακτηριστικά και για το ποιόν των οποίων έχουν περιορισμένη πληροφόρηση. Χρειαζόμαστε λοιπόν κάτι πέρα από τη θεωρία της επιλογής βάσει γενετικής συγγένειας για να εξηγήσουμε τη συνεργασία σε όλο της το εύρος. Η βασική θεωρία που συμβαδίζει με την υπόθεση των ορθολογικών εγωιστών και προσφέρει μια ευρεία βάση εξήγησης για τη συνεργατική συμπεριφορά είναι αυτή της *άμεσης αμοιβαιότητας ή του αμοιβαίου αλτρουισμού (direct reciprocity, reciprocal altruism)*. (Trivers, 1971; Axelrod & Hamilton, 1981; Nowak & Highfield, 2011, ch. 1).

Σύμφωνα με αυτή, τα άτομα ανταποδίδουν τη συνεργατική συμπεριφορά των άλλων ατόμων και είναι διατεθειμένα ακόμη και να αναλάβουν κόστη για να τιμωρήσουν ή να επιβραβεύσουν τους άλλους, αρκεί να προσδοκούν σε ένα άμεσο ή έμμεσο υλικό όφελος. Είναι δηλαδή απαραίτητο να υπάρχει ένας επαρκώς μεγάλος μελλοντικός ορίζοντας για τη συνέχιση

της συνεργασίας που να καθιστά τις αλληλεπιδράσεις μεταξύ των ίδιων ατόμων επαναλαμβανόμενες και τη συνεργασία κερδοφόρα. Οι μακροχρόνιες σχέσεις καθιστούν ασύμφορη την τακτική της αθέτησης των υποχρεώσεων, εφόσον μια τέτοια στρατηγική θα προκαλέσει απόσυρση της συνεργασίας από τους υπόλοιπους συναλλασσόμενους και θα οδηγήσει σε μελλοντική απώλεια υλικών απολαβών. Η επιβολή τιμωρίας με κόστος όταν τα άτομα συμπεριφέρονται σύμφωνα με την άμεση αμοιβαιότητα έχει επομένως νόημα σε ένα ορθολογικά εγωιστικό πλαίσιο, εφόσον συντηρεί τη συνεργασία και τα μελλοντικά οφέλη.

Το χαρακτηριστικό παράδειγμα της συνεργασίας σε αυτή τη βάση έχει μοντελοποιηθεί στη θεωρία παιγνίων σε μια στρατηγική που είναι γνωστή στη βιβλιογραφία ως Tit for Tat (Axelrod & Hamilton, 1981; Axelrod, 1984). Η στρατηγική αυτή έχει προσομοιωθεί σε διμερείς αλληλεπιδράσεις και έχει βρεθεί ότι αποτελεί την πιο αποτελεσματική επιλογή σε παίγνια όπως το δίλημμα του φυλακισμένου. Τα άτομα που την ακολουθούν ξεκινούν το παιχνίδι με μια συνεργατική επιλογή και στη συνέχεια σε κάθε επόμενη επανάληψη του παιχνιδιού μιμούνται την επιλογή που έκανε ο άλλος παίκτης στον προηγούμενο γύρο. Αν ο παίκτης συνεργάστηκε τότε συνεργάζονται και αυτά, ενώ αν πρόδωσε τη συνεργασία, προδίδουν και αυτά με τη σειρά τους. Αυτή η στρατηγική αποδίδει τα μεγαλύτερα οφέλη και άρα είναι σταθερή και από την εξελικτική σκοπιά. Ωστόσο, όταν αυτή η μορφή ανταπόδοσης που βασίζεται στο μακροχρόνιο ορίζοντα των επαναλήψεων εφαρμοστεί σε μεγάλες ομάδες ατόμων (πάνω από 8 μέλη), η σταθερότητα αυτή κλονίζεται (Boyd & Richerson, 1988). Ο λόγος είναι ότι το αντίκτυπο της μεμονωμένης συμπεριφοράς του καθενός παίκτη δεν είναι πια τόσο άμεσο για τους υπόλοιπους. Έτσι, η απόσυρση της συνεργασίας από ένα μεμονωμένο άτομο δεν πλήττει τόσο πολύ τις απολαβές των υπολοίπων και, στην περίπτωση που ειδωθεί σαν μια έμμεση μορφή τιμωρίας, δε μπορεί να στοχεύσει τα μη συνεργατικά μέλη. Η εξάρτηση λοιπόν μεταξύ των παικτών παύει να είναι άμεση και το αποτέλεσμα είναι η κατάρρευση της συνεργασίας μετά από μερικές επαναλήψεις του παιχνιδιού.

1.3 Έμμεση αμοιβαιότητα

Έτσι λοιπόν, το ερώτημα για τους μηχανισμούς που στηρίζουν τη συνεργασία και την επιβολή τιμωριών για τη διατήρησή της φαίνεται να χρειάζεται κάποιες πρόσθετες εξηγήσεις. Μια τέτοια προσπάθεια έχει γίνει με τη θεωρία της *έμμεσης αμοιβαιότητας (indirect reciprocity)* (Nowak &

Sigmund, 1998).

Η θεώρηση αυτή περιλαμβάνει την προσθήκη κάποιων επιπλέον παραμέτρων οι οποίες είναι πολύ σημαντικές στις συνεργατικές αλληλεπιδράσεις. Οι παράμετροι αυτές αφορούν το ρόλο που παίζει στις συνεργατικές επιλογές η δημιουργία φήμης για τα άτομα. Ένα άτομο δηλαδή μπορεί να προβαίνει σε μια συμπεριφορά που δεν έχει κάποιο άμεσο ή έμμεσο υλικό όφελος προκειμένου σε άλλες αλληλεπιδράσεις με άλλα άτομα στο μέλλον να επωφεληθεί από το καλό όνομα που έχει δημιουργήσει. Αν υπάρχει επαρκής πληροφόρηση για την ταυτότητα και το ιστορικό των ατόμων σε μια ομάδα, η γνώση ότι ένα άτομο είναι γενικά συνεργατικό μπορεί να ωθήσει τα άλλα άτομα να συνεργαστούν μαζί του όταν το συναντήσουν. Στο πλαίσιο αυτό, με την επιβολή τιμωριών ένα άτομο μπορεί να δημιουργήσει τη φήμη ότι δεν ανέχεται τις παραβιάσεις των συνεργατικών προτύπων και τις τιμωρεί ακόμα κι αν αυτό του κοστίζει. Η απειλή αυτή μπορεί να συμβάλλει στη συμμόρφωση των άλλων ατόμων με συνεργατικά πρότυπα όταν αλληλεπιδρούν μαζί του. Έτσι, τελικά στο βαθύ μέλλον υπάρχει και πάλι ένα υλικό όφελος για το άτομο. Στο ίδιο μήκος κύματος κινείται και η θεωρία της *κοστοβόρου σηματοδότησης (costly signaling)* (Gintis et al., 2001; McAndrew, 2002). Σύμφωνα με αυτή τα άτομα μπορεί να παρουσιάζουν συνεργατικές συμπεριφορές σε ένα τομέα αλληλεπιδράσεων προκειμένου να κινητοποιήσουν τη συνεργατική συμπεριφορά των άλλων μελών της ομάδας απέναντί τους σε άλλους τομείς αλληλεπίδρασης.

Οι προηγούμενες θεωρίες ενσωματώνουν το στοιχείο της φήμης και της πληροφόρησης το οποίο είναι αναμφίβολα εξαιρετικά σημαντικό στις ανθρώπινες αλληλεπιδράσεις. Οι ισχυρισμοί τους εξακολουθούν να καλύπτονται από τις υποθέσεις της ορθολογικής εγωιστικής μεγιστοποίησης και έχουν ισχύ και στο εξελικτικό επίπεδο. Ωστόσο, απαραίτητη προϋπόθεση αποτελεί και πάλι ο μακροχρόνιος ορίζοντας των σχέσεων και η ύπαρξη επαρκούς πληροφόρησης για τα μέλη της ομάδας, αλλιώς και πάλι η συνεργασία δε μπορεί να εξηγηθεί επαρκώς σε όλο της το εύρος.

1.4 Ισχυρή αμοιβαιότητα και αλτρουιστική τιμωρία

Οι θεωρίες που περιγράφηκαν ανωτέρω καλύπτουν ένα μεγάλο μέρος πειραματικών ευρημάτων στη συμπεριφορική θεωρία παιγνίων, όταν τα άτομα εμπλέκονται σε επαναλαμβανόμενες αλληλεπιδράσεις. Η μεγιστοποίηση του προσωπικού υλικού συμφέροντος, άσχετα από τις

απολαβές των άλλων, αποτελεί επαρκές κίνητρο για να εξηγήσει τις συνεργατικές συμπεριφορές όταν τα άτομα αλληλεπιδρούν με συγγενείς ή με άλλα άτομα τα οποία πρόκειται να ξανασυναντήσουν στο μέλλον ή με τα οποία έχουν εγκαθιδρύσει μακροχρόνιες σχέσεις.

Ωστόσο, στην καθημερινή ζωή παρατηρείται συνεργατική συμπεριφορά σε μια κατηγορία αλληλεπιδράσεων που δε μπορεί να ενταχθεί στο επεξηγηματικό πλαίσιο των παραπάνω θεωριών. Πρόκειται για τις αλληλεπιδράσεις μεταξύ ξένων, οι οποίες δεν έχουν ένα μακροχρόνιο ορίζοντα επαναλήψεων, καθώς και για τις αλληλεπιδράσεις μεταξύ πολλών ατόμων. Μια σειρά πειραματικών ευρημάτων στην κατεύθυνση αυτή παρουσιάζει επίσης ασυμβατότητα με τις θεωρίες αυτές. Το βασικό εύρημα που δεν καλύπτεται είναι το γιατί τα άτομα επιδεικνύουν συνεργατικές συμπεριφορές σε *στιγμιαίες αλληλεπιδράσεις μιας ευκαιρίας (one-shot)* και σε ανώνυμες αλληλεπιδράσεις. Στις ανώνυμες αλληλεπιδράσεις τα άτομα δεν έχουν καμία πληροφορία για το άτομο με το οποίο αλληλεπιδρούν και δε γνωρίζουν τίποτα για το ιστορικό του. Προσομοιάζει δηλαδή την αλληλεπίδραση δύο απολύτως ξένων μεταξύ τους ατόμων, κάτι που συμβαίνει συχνά στην καθημερινότητα. Όταν η αλληλεπίδραση είναι επιπλέον και στιγμιαίες (*one-shot*), τότε μπορούμε να αποκλείσουμε οποιαδήποτε επίδραση της φήμης (*reputation*) στα αποτελέσματα, εφόσον τα άτομα γνωρίζουν ότι δε θα συναντηθούν ξανά.

Σκοπός των πειραμάτων που διεξάγονται κάτω από αυτούς τους όρους είναι να απομονώσουν τις εγγενείς συμπεριφορικές αποκρίσεις των ατόμων σε καταστάσεις όπου καλούνται να κάνουν συνεργατικές επιλογές, αφήνοντας έξω όλους τους άλλους παράγοντες που μπορεί να τις επηρεάζουν. Μπορούμε έτσι να εντοπίσουμε εγγενείς γνωσιακούς μηχανισμούς που πυροδοτούν αυτές τις αποκρίσεις. Ο πειραματικός χειρισμός είναι απαραίτητος γιατί είναι η μόνη μέθοδος με την οποία μπορούμε να απομονώσουμε κίνητρα πέρα από τα υλικά εγωιστικά που επικαλούνται η έμμεση και η άμεση αμοιβαιότητα. Στην πραγματικό κόσμο είναι αδύνατο να απομονωθούν και να διαχωριστούν αυτές οι συνιστώσες, για αυτό και είναι αναγκαία η χρήση πειραμάτων σχεδιασμένων στο εργαστήριο, όπου είναι δυνατός ο έλεγχος και ο διαχωρισμός των παραμέτρων.

Στις περιπτώσεις των ανώνυμων one-shot αλληλεπιδράσεων μεταξύ πολλών ατόμων, καμία από τις προηγούμενες θεωρίες δεν είναι σε θέση να παρέχει ένα επαρκές ερμηνευτικό πλαίσιο για το υπόβαθρο κινήτρων που ωθεί τα άτομα σε συμπεριφορές συνεργασίας. Τα άτομα κάνουν επιλογές που αντιβαίνουν το κίνητρο της μεγιστοποίησης του προσωπικού συμφέροντος με στενούς οικονομικούς όρους: στις αποφάσεις τους φαίνεται να παίζει ρόλο ένα πιο ευρύ πλαίσιο

προτιμήσεων από τις αυστηρά οικονομικές. Οι πεποιθήσεις, οι αξίες, τα πρότυπα και ένα σύνολο από εσωτερικοποιημένους κανόνες κοινωνικής συμπεριφοράς (*social norms*) δείχνουν να επηρεάζουν την συναισθηματική και ψυχολογική τους πραγματικότητα και δημιουργούν ένα πλέγμα κινήτρων που έχει σαν επίκεντρο τη σχέση τους με τα άλλα άτομα. Η συμπεριφορά τους δηλαδή σε μεγάλο βαθμό αφορά τη σχέση τους με τους άλλους (*other-regarding behavior*) και φαίνεται ότι περιλαμβάνει μια σειρά *κοινωνικών προτιμήσεων (social preferences)* (Falk & Fischbacher, 2006; Bolton & Ockenfels, 2000). Η συμπεριφορά τους έχει έντονα στοιχεία αμοιβαιότητας, ακόμη και όταν τα άτομα πρέπει να επωμιστούν προσωπικό κόστος προκειμένου να την επιδείξουν.

Τα πειραματικά αυτά ευρήματα δημιουργούν ένα ερώτημα για το κατά πόσο οι θεωρίες της άμεσης και έμμεσης αμοιβαιότητας αρκούν για να προσφέρουν ένα ικανοποιητικό θεωρητικό πλαίσιο για την εξέλιξη της συνεργασίας. Οι συνεργατικές συμπεριφορές που παρουσιάζουν οι άνθρωποι στην καθημερινότητα δείχνουν να έχουν μια εγγενή συνιστώσα, η εξελικτική ύπαρξη της οποίας δε θεμελιώνεται επαρκώς από τα κίνητρα των ανωτέρω θεωριών. Ένα πλήθος πειραμάτων στη συμπεριφορική θεωρία παιγνίων αποκαλύπτουν μια εγγενή ροπή των ατόμων να εκτιμούν και να επιθυμούν να επιβραβεύσουν τις συμπεριφορές που αντιλαμβάνονται ως δίκαιες και να επικρίνουν και να επιθυμούν να τιμωρήσουν όσες αντιλαμβάνονται ως άδικες. Τα άτομα δηλαδή τείνουν να επιδεικνύουν συμπεριφορές δαπανηρής επιβράβευσης και δαπανηρής τιμωρίας *χωρίς να περιμένουν κάποιο βραχυπρόθεσμο ή μακροπρόθεσμο όφελος*. Το στοιχείο αυτό αποτελεί τη βάση της θεωρίας της *ισχυρής αμοιβαιότητας (strong reciprocity theory)* (Fehr et al., 2002; Gintis, 2000; Gintis et al., 2005; Henrich et al., 2004). Σύμφωνα με αυτή, τα άτομα έχουν τη προδιάθεση να:

- θυσιάσουν πόρους για να ανταμείψουν μια δίκαιη, σύμφωνη με τους κοινωνικούς κανόνες συμπεριφορά (θετική αμοιβαιότητα, *αλτρουιστική ή δαπανηρή επιβράβευση, altruistic or costly rewarding*)
- θυσιάσουν πόρους για να τιμωρήσουν μια άδικη συμπεριφορά που παραβιάζει τους κοινωνικούς κανόνες (αρνητική αμοιβαιότητα*, αλτρουιστική ή δαπανηρή τιμωρία, altruistic or costly punishment*) (Gintis, 2009).

Τα άτομα που παρουσιάζουν συμπεριφορές ισχυρής αμοιβαιότητας αναλαμβάνουν το κόστος της τιμωρίας ακόμη και αν δεν αναμένουν κανένα ατομικό οικονομικό όφελος από τις πράξεις τους. Σε αντίθεση με αυτό, τα άτομα που συμπεριφέρονται όπως προβλέπει η θεωρία της άμεσης

αμοιβαιότητας, επιβραβεύουν και τιμωρούν μόνο αν αυτό συνάδει με το μακροπρόθεσμο ατομικό τους συμφέρον. Έτσι η ισχυρή αμοιβαιότητα αποτελεί ένα ισχυρό κίνητρο για συνεργασία ακόμη και σε μη επαναλαμβανόμενες αλληλεπιδράσεις και όταν δεν υπάρχουν οφέλη από τη φήμη, επειδή τα άτομα που επιδεικνύουν αυτή τη συμπεριφορά θα επιβραβεύσουν αυτούς που συνεργάζονται και θα τιμωρήσουν αυτούς που προδίδουν τη συνεργασία.

Κεφάλαιο 2: Η αλτρουιστική τιμωρία: ενδείξεις και ευρήματα από πειράματα και μελέτες

Από τη σκοπιά της θεωρίας της ισχυρής αμοιβαιότητας ένας βασικός μηχανισμός που παίζει ρόλο στην ανάπτυξη και την σταθεροποίηση της συνεργασίας και των κοινωνικών κανόνων είναι η αλτρουιστική ή δαπανηρή τιμωρία, δηλαδή η τιμωρία που επιβάλλει ένα άτομο επωμιζόμενο ένα οικονομικό κόστος, χωρίς να περιμένει κάποιο οικονομικό όφελος ή όφελος γοήτρου. Κίνητρο για αυτή την πράξη δεν είναι επομένως η εγωιστική μεγιστοποίηση του οικονομικού κέρδους ή η απόκτηση επωφελούς φήμης, αλλά μια εγγενής προδιάθεση του ατόμου να απορρίπτει και να τιμωρεί τις συμπεριφορές που κρίνει ως κοινωνικά παραβατικές ή άδικες. Η εγγενής αυτή προδιάθεση βασίζεται στην ύπαρξη νοητικών μηχανισμών που έχουν διαμορφωθεί στην πορεία της ανθρώπινης εξέλιξης.

Στην καθημερινή ζωή παρατηρούμε ένα πλήθος περιπτώσεων όπου τα άτομα αναλαμβάνουν κάποιο κόστος προκειμένου να επιβραβεύσουν ή να τιμωρήσουν μια συμπεριφορά που δε συμβαδίζει με τους κοινωνικούς κανόνες κοινωνικής συμπεριφοράς. Πολλές από αυτές τις περιπτώσεις έχουν καταγραφεί σε ανθρωπολογικές και εθνογραφικές μελέτες (Henrich et al., 2001; Henrich et al., 2006; Fessler, 2002; Boyd & Richerson, 2005). Προκειμένου ωστόσο να διερευνηθούν τα κίνητρα των πράξεων αυτών η μελέτη σε συνθήκες πραγματικής ζωής δε μπορεί να βοηθήσει γιατί είναι ουσιαστικά αδύνατο στις συνθήκες αυτές να διαχωρίσουμε μεταξύ των κινήτρων που προκαλούν την τιμωρητική συμπεριφορά. Δε μπορούμε δηλαδή να γνωρίζουμε αν αυτό που ωθεί το άτομο να τιμωρήσει είναι η γενετική συγγένεια, η αναμονή κάποιου μελλοντικού οφέλους σε μια επόμενη συνάντηση, η δημιουργία φήμης ως ατόμου που δεν ανέχεται την αδικία ή μια εσωτερική έμφυτη τάση.

Οι τρεις πρώτες από τις παραπάνω κατηγορίες κινήτρων μπορούν να ενταχθούν στο πλαίσιο μιας εγωιστικής μεγιστοποιητικής συμπεριφοράς και καλύπτονται από τις θεωρίες της επιλογής βάσει συγγένειας και τις θεωρίες της άμεσης και έμμεσης αμοιβαιότητας αντίστοιχα. Για να μπορέσουμε να εντοπίσουμε εάν υφίσταται η τέταρτη κατηγορία και επομένως η αλτρουιστική τιμωρία, πρέπει να σχεδιάσουμε πειραματικές διατάξεις στο εργαστήριο που θα απομονώνουν τους άλλους παράγοντες. Χρειάζονται δηλαδή συνθήκες όπου τα άτομα δεν είναι συγγενείς, γνωρίζουν ότι αποκλείεται η περίπτωση μιας επαναληπτικής συνάντησης και δε γίνεται να υπάρχει πληροφορία για την ταυτότητα των ατόμων που αλληλεπιδρούν, οι αλληλεπιδράσεις

δηλαδή είναι ανώνυμες και one-shot . Οι συμμετέχοντες πρέπει να είναι πλήρως κατατοπισμένοι για αυτές τις συνθήκες και τα οικονομικά οφέλη από το πείραμα να είναι πραγματικά. Τα αποτελέσματα τέτοιων πειραμάτων μπορούν να μας προσφέρουν ενδείξεις για την ύπαρξη και τη λειτουργία της αλτρουιστικής τιμωρίας, καθώς και για τις επιδράσεις που μπορεί να έχει σε αυτή τη συμπεριφορά η πιθανότητα μελλοντικών συναντήσεων και δημιουργίας φήμης. Μας βοηθούν επομένως να διαχωρίσουμε τα κίνητρα και να δούμε πώς διαπλέκονται μεταξύ τους οδηγώντας στις παρατηρούμενες συνεργατικές συμπεριφορές.

Παρακάτω παρατίθεται μια σειρά τέτοιων πειραματικών ευρημάτων που παρέχουν ενδείξεις για την ύπαρξη αυτής της εσωτερικής τάσης των ανθρώπων να τιμωρούν τις μη συνεργατικές συμπεριφορές, ακόμα και όταν αυτό έχει κόστος και όχι υλικό όφελος για τους ίδιους. Η παρουσία αυτής της εγγενούς συνιστώσας για την αλτρουιστική συμπεριφορά δεν καλύπτεται από τις υποθέσεις των εγωιστικών κινήτρων της άμεσης και της έμμεσης αμοιβαιότητας, ούτε βέβαια από την επιλογή βάσει συγγένειας, αφού αφορά ξένα μεταξύ τους άτομα. Ενισχύει λοιπόν την υπόθεση για την ύπαρξη μιας επιπλέον κατηγορίας κινήτρων στους ανθρώπους που αφορούν τις σχέσεις τους με τους άλλους και ονομάζονται κοινωνικές προτιμήσεις. Επιπλέον ενδείξεις για τα κίνητρα αυτά και τους γνωσιακούς μηχανισμούς με τους οποίους εμπλέκονται στη διαδικασία της λήψης αποφάσεων έρχονται από νευροαπεικονιστικές μελέτες. Εξελικτικά μοντέλα επίσης δείχνουν ότι η συμπεριφορά που προκαλούν τα κίνητρα αυτά μπορεί να χαρακτηρίζεται από εξελικτική σταθερότητα. Το σύνολο αυτών των μελετών που παρουσιάζονται παρακάτω μας βοηθά να βγάλουμε συμπεράσματα για την ύπαρξη, τη φύση και τα όρια της αλτρουιστικής τιμωρίας καθώς και για το ρόλο και την επίδρασή της στην εδραίωση της συνεργασίας και των κοινωνικών κανόνων.

2.1 Συμπεριφορικά πειράματα

Στην καθημερινή ζωή παρατηρείται πολύ συχνά η τάση των ανθρώπων να ζητούν εκδίκηση σαν απάντηση σε άδικες και επιβλαβείς πράξεις. Στις διάφορες αλληλεπιδράσεις μεταξύ τους τα άτομα εκτιμούν τις απολαβές των συμμετεχόντων στην αλληλεπίδραση, καθώς και των προθέσεων που αποκαλύπτουν οι επιλογές τους. Η εκτίμηση περιλαμβάνει έναν υπολογισμό του οφέλους που αποκομίζει ο κάθε συμμετέχων καθώς και μια αξιολογική κρίση για το πόσο δίκαιο ήταν αυτό το κέρδος και πόσο δίκαια συμπεριφέρθηκε ο καθένας προκειμένου να το αποκτήσει.

Η κρίση αυτή περί δικαιοσύνης εξαρτάται από τους θεσμικούς κανόνες και τις συνθήκες της συγκεκριμένης αλληλεπίδρασης αλλά και από γενικούς, άγραφους κανόνες κοινωνικής συμπεριφοράς (Fehr & Schmidt, 1999). Όταν αυτοί οι κανόνες παραβιάζονται, οι άνθρωποι παρουσιάζουν την τάση να τιμωρούν τους παραβάτες και να εξισώνουν τις απολαβές. (Dawes et al., 2007). Όπως αναφέρθηκε προηγουμένως, προκειμένου να διαχωρίσουμε τα κίνητρα που βρίσκονται πίσω από τέτοιες πράξεις, είναι χρήσιμο να μελετήσουμε τη συμπεριφορά κάτω από πειραματικά σχεδιασμένες συνθήκες. Οι πειραματικοί σχεδιασμοί της συμπεριφορικής θεωρίας παιγνίων (*behavioral game theory*) χρησιμοποιούνται συχνά για το σκοπό αυτό (Camerer, 2003).

2.1.1 Ultimatum game (one-shot)

Μια χαρακτηριστική απεικόνιση της τάσης των ατόμων να τιμωρούν τις άδικες κατανομές απολαβών και να προσπαθούν να εξισώσουν την κατανομή υλικών οφελών προσφέρουν τα ευρήματα του παιχνιδιού του τελεσίγραφου (*ultimatum game*). Στο ultimatum game (Güth et al., 1982; Camerer & Thaler, 1995) παίρνουν μέρος δύο συμμετέχοντες οι οποίοι πρέπει να συμφωνήσουν για το μοίρασμα ενός σταθερού χρηματικού ποσού. Ο πρώτος συμμετέχων έχει το ρόλο του του Προτείνοντα (*Proposer*) και ο δεύτερος το ρόλο του Δέκτη (*Responder*). Ο Προτείνων μπορεί να κάνει ακριβώς μία πρόταση για το πώς θα μοιραστεί το ποσό. Στη συνέχεια ο Δέκτης μπορεί να να αποδεχτεί ή να απορρίψει την πρόταση. Οι συμμετέχοντες δεν έχουν δυνατότητα διαπραγμάτευσης και σε περίπτωση απόρριψης κανείς δεν κερδίζει τίποτα. Η αλληλεπίδραση μεταξύ των ατόμων είναι *one-shot*, το παιχνίδι δηλαδή δεν επαναλαμβάνεται με τον ίδιο συμπαίκτη, και ανώνυμη, τα άτομα δηλαδή δεν έχουν πληροφορία για την ταυτότητα του συμπαίκτη τους. Σε περίπτωση αποδοχής μοιράζονται το ποσό σύμφωνα με την πρόταση. Π.χ. αν το ποσό είναι 100 €, ο Προτείνων μπορεί να κάνει μια πρόταση να δώσει 20€ στο Δέκτη και να κρατήσει 80€ ο ίδιος. Αν ο Δέκτης δεχτεί την πρόταση παίρνουν ο καθένας το αντίστοιχο ποσό, ενώ αν την απορρίψει κανείς τους δεν παίρνει τίποτα.

Ένα ισχυρό αποτέλεσμα σε αυτό το πείραμα είναι ότι προτάσεις που δίνουν στο Δέκτη μερίδιο κάτω από 25% του διαθέσιμου ποσού απορρίπτονται με πολύ μεγάλη πιθανότητα. Αυτό δείχνει ότι οι Δέκτες δε συμπεριφέρονται με τρόπο που θα μεγιστοποιούσε το προσωπικό τους υλικό συμφέρον. Σε αυτή την περίπτωση οποιαδήποτε προσφορά πάνω από το μηδέν θα γινόταν αποδεκτή από το Δέκτη, εφόσον και το ελάχιστο κέρδος είναι προτιμότερο από το μηδενικό. Απ' ό,τι φαίνεται όμως οι Δέκτες στην κρίση τους περιλαμβάνουν και τη σύγκριση με το κέρδος του

άλλου ατόμου και το πόσο δίκαιη τους φαίνεται η μοιρασιά. Η τυπικά δίκαιη μοιρασιά θεωρείται η ισόποση, δηλαδή να μοιραστεί το ποσό στη μέση. Αν κρίνουν ότι η μοιρασιά είναι υπερβολικά άδικη, προτιμούν να θυσιάσουν το προσωπικό τους υλικό κέρδος προκειμένου να τιμωρήσουν με αυτό τον τρόπο τον Προτείνοντα, στερώντας του τη δική του απολαβή. Η συμπεριφορά αυτή είναι σύμφωνη με τη θεωρία της ισχυρής αμοιβαιότητας, εφόσον το άτομο θυσιάζει πόρους χωρίς να περιμένει κάποια άμεσο ή μελλοντικό αντίκρυσμα για αυτή τη θυσία, αφού η αλληλεπίδραση είναι one-shot και ανώνυμη. Αποτελεί δηλαδή ένα παράδειγμα αλτρουιστικής τιμωρίας. Οι Δέκτες δεν ξανασυναντούν τον ίδιο Προτείνοντα σε επόμενες επαναλήψεις, ωστόσο η απόρριψη των χαμηλών προσφορών έχει επίδραση στη συμπεριφορά των Προτεινόντων. Όταν η προσφορά τους απορρίπτεται σε κάποιο γύρο, στις νέες προσφορές που κάνουν στους επόμενους Δέκτες που συναντούν παρατηρείται αύξηση της τάξης του 7% του διαθέσιμου ποσού (Fehr & Fischbacher, 2003). Η έμμεση δηλαδή τιμωρία τους μέσω της απόρριψης τους ωθεί να ακολουθήσουν το δίκαιο πρότυπο κατανομής και να κάνουν λιγότερο άνισες προσφορές στους επόμενους γύρους.

Τα αποτελέσματα αυτά έχουν παρατηρηθεί σε πολλές πειραματικές μελέτες του ultimatum game και δεν αλλάζουν ακόμα και σε περιπτώσεις που οι διαθέσιμες χρηματικές απολαβές από το παιχνίδι είναι μεγάλες (πχ. ποσά που αντιπροσωπεύουν μέσα εισοδήματα τριών μηνών, Cameron, 1999; Slonim & Roth, 1998). Αντικατοπτρίζουν επομένως μια πραγματικά ισχυρή και γενική τάση των ατόμων και δεν οφείλονται στο ότι τα άτομα δε θεωρούν το χρηματικό διακύβευμα σημαντικό. Επίσης τα αποτελέσματα αυτά έχουν και διαπολιτισμική ισχύ, εφόσον ανευρίσκονται σε ένα μεγάλο αριθμό διαφορετικών χωρών όπου έχουν διεξαχθεί μελέτες με one-shot ultimatum games, όπως οι ΗΠΑ, η Ιαπωνία, η Ρωσία, το Ισραήλ, η Ινδονησία και πολλές Ευρωπαϊκές χώρες (Roth et al., 1991; Henrich et al., 2001). Έχει μάλιστα παρατηρηθεί ότι όσο πιο διαδεδομένοι είναι οι μηχανισμοί της αγοράς σε μια κοινωνία, το κατώφλι απόρριψης και η μέση προσφορά ανεβαίνουν (Henrich, 2000).

Το ultimatum game αποτελεί μια ευρέως μελετημένη ένδειξη των αρνητικών συναισθημάτων που προκαλεί η αίσθηση ότι κάποιος αδικείται και της επιθυμίας να δράσει προκειμένου να διορθώσει αυτή την αδικία, τιμωρώντας το άτομο που την προκάλεσε. Το ότι αυτά τα συναισθήματα αποτελούν μια γνωστή και αναγνωρισμένη ψυχολογική πραγματικότητα είναι εμφανές από το γεγονός ότι οι Προτείνοντες αναμένουν αυτή την αντίδραση από την πλευρά των Δεκτών και για αυτό σε μεγάλο ποσοστό κάνουν προτάσεις που πλησιάζουν τη

δίκαιη μοιρασιά. Σε πειράματα ultimatum που έχουν διεξαχθεί στη Δυτική Ευρώπη και τη Βόρεια Αμερική, η μέση προσφορά ήταν στο επίπεδο του 30-40% του διαθέσιμου ποσού, με μια τυπική τιμή κοντά στο 50-50 (Camerer, 2003). Αντίθετα, στο παιχνίδι του δικτάτορα (dictator's game) όπου ο Δέκτης δεν έχει τη δυνατότητα να απορρίψει την προσφορά και είναι υποχρεωμένος να δεχτεί ο,τι του προσφέρει ο Προτείνων-δικτάτορας, η μέση προσφορά είναι πολύ χαμηλότερη απ' ό,τι στο ultimatum. Ο ρόλος λοιπόν της αλτρουιστικής τιμωρίας στην προώθηση συνεργατικών συμπεριφορών μπορεί να διαφανεί ήδη μέσα από αυτό το απλό παιχνίδι. (Forsythe et al., 1994; Hoffman et al., 1994).

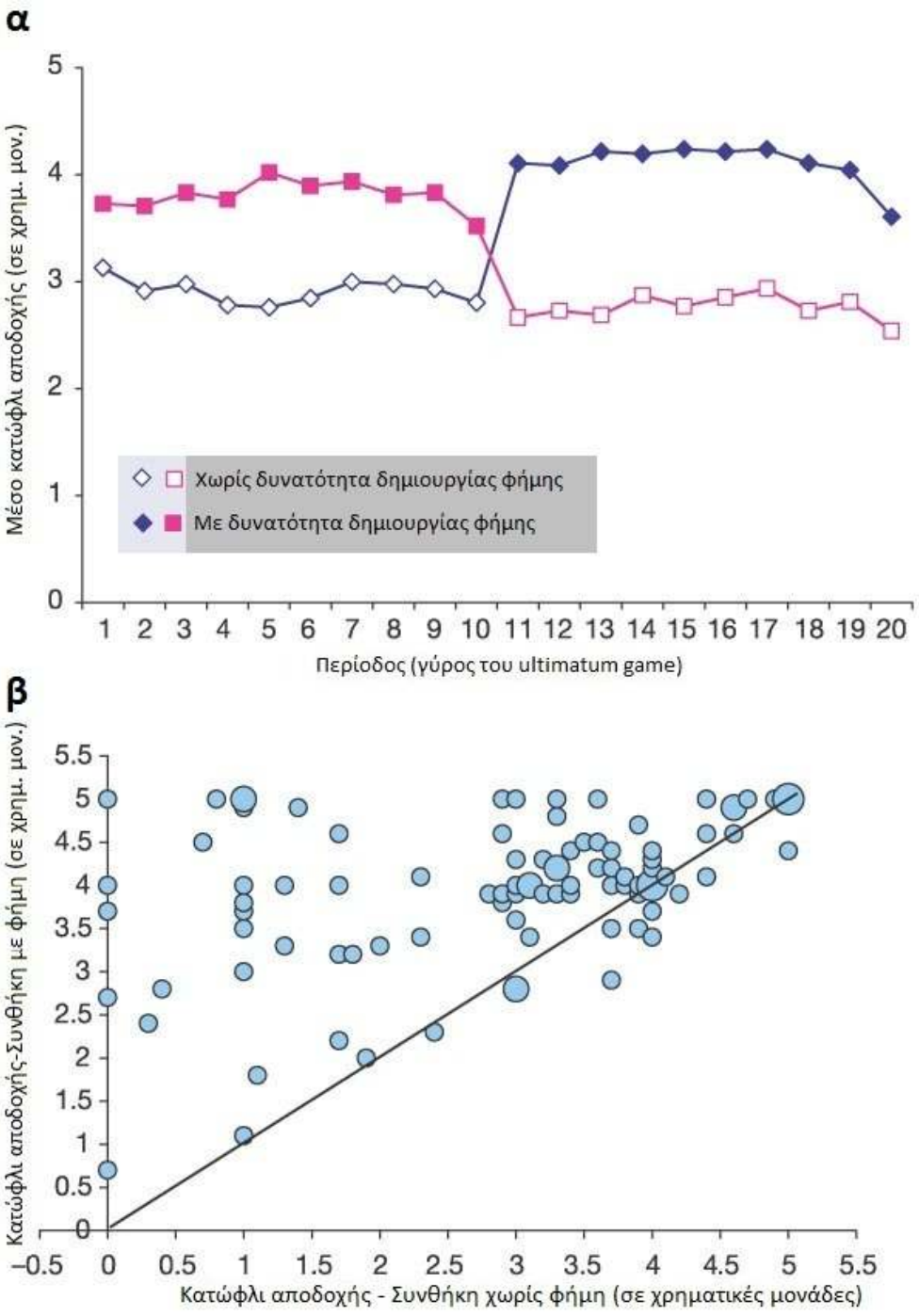
Θα μπορούσε αυτή συμπεριφορά να δικαιολογηθεί από τα κίνητρα των θεωριών της άμεσης ή της έμμεσης αμοιβαιότητας; Εφόσον η αλληλεπίδραση στο ultimatum είναι one-shot και ανώνυμη η απάντηση είναι αρνητική. Θα μπορούσε ωστόσο κανείς να ισχυριστεί ότι τα άτομα δε μπορούν να αντιληφθούν το γεγονός ότι η φήμη (*reputation*) ή μια νέα συνάντηση με το ίδιο άτομο δεν παίζουν κανένα ρόλο σε αυτές τις συνθήκες και συμπεριφέρονται όπως έχουν συνηθίσει στην καθημερινή τους ζωή. Στην καθημερινότητα πολλές φορές περιμένουμε μια επόμενη συνάντηση με το ίδιο άτομο και γενικά θέλουμε να χτίσουμε μια φήμη που θα μας ωφελήσει όταν συναντούμε νέα άτομα. Θα μπορούσε επομένως να παρατηρηθεί στην πειραματική συμπεριφορά ένα φαινόμενο μεταφοράς κάποιων ευρετικών (*heuristics*), κάποιων προτύπων ή στρατηγικών επιτυχημένης συμπεριφοράς που έχουμε μάθει να χρησιμοποιούμε στις συνθήκες της καθημερινότητας. Κάτι τέτοιο έχει ελεγχθεί πειραματικά και έχει βρεθεί ότι δεν ισχύει. Οι συμμετέχοντες είναι σε θέση να διαχωρίσουν τις συνθήκες και κατανοούν ποια είναι η σημασία των επαναλαμβανόμενων αλληλεπιδράσεων και της φήμης.

2.1.2 Ultimatum (με δημιουργία φήμης)

Προκειμένου να ελεγχθεί αυτή η υπόθεση για το ultimatum διεξήχθη μια παραλλαγή του πειράματος με δύο συνθήκες: μια συνθήκη φήμης και μια συνθήκη βάσης (Fehr & Fischbacher, 2003). Και στις δύο συνθήκες διεξάγονται δέκα γύροι και 10 χρηματικές μονάδες πρέπει να μοιραστούν σε κάθε γύρο, σε καθέναν από τους οποίους κάθε Προτείνων συναντά ένα νέο Δέκτη. Στη συνθήκη φήμης οι Προτείνοντες πληροφορούνται σχετικά με το ιστορικό απορρίψεων του Δέκτη που συναντούν, ενώ στη συνθήκη βάσης δεν παίρνουν καμία πληροφορία. Αυτό σημαίνει ότι οι Δέκτες στη συνθήκη φήμης μπορούν να αποκτήσουν μια φήμη σκληρού διαπραγματευτή απορρίπτοντας ακόμη και υψηλές προσφορές. Υπάρχει δηλαδή

η δυνατότητα για δημιουργία φήμης (*reputation formation*). Το βραχυπρόθεσμο κόστος της απόρριψης μιας πρότασης μπορεί να αντισταθμιστεί από το μακροπρόθεσμο όφελος μιας καλής φήμης που θα οδηγήσει τους επόμενους προτείνοντες να κάνουν καλύτερες προσφορές. Στη συνθήκη βάσης μια απόρριψη πρότασης δεν είναι δυνατό να αποφέρει αυτό το όφελος, επομένως αν οι συμμετέχοντες αντιλαμβάνονται τη διαφορά ανάμεσα στις δύο συνθήκες θα πρέπει να έχουν χαμηλότερα κατώφλια απόρριψης στη συνθήκη βάσης.

Τα αποτελέσματα δείχνουν ότι όταν τα άτομα παίζουν πρώτα στη συνθήκη βάσης και μετά στη συνθήκη της φήμης, το μέσο κατώφλι αποδοχής ανεβαίνει από τις 3 στις 4 χρηματικές μονάδες. Η συντριπτική πλειοψηφία των Δεκτών (84%) αυξάνει το κατώφλι ενώ η εναπομένουσα μειοψηφία το διατηρεί περίπου σταθερό (Σχήμα 1α,1β). Η αύξηση του κατωφλιού οδηγεί τους Προτείνοντες να αυξήσουν τις προσφορές τους. Αντίστοιχα, όταν τα άτομα παίζουν πρώτα στη συνθήκη φήμης και μετά στη συνθήκη βάσης το κατώφλι κατεβαίνει. Τα αποτελέσματα αυτά δείχνουν καθαρά ότι τα άτομα είναι σε θέση να διαφοροποιήσουν τις δύο συνθήκες και να διαχωρίσουν τις συνέπειες που συνεπάγεται η απόρριψη της προσφοράς στην καθεμία. Ο ρόλος και η ισχύς της αλτρουιστικής τιμωρίας είναι ακόμη πιο μεγάλη στη συνθήκη όπου μπορεί να υπάρξει διαμόρφωση φήμης και τα άτομα γνωρίζουν πώς να το εκμεταλλευτούν αυτό προκειμένου να εκμαιεύσουν πιο συνεργατικές συμπεριφορές από τους αντιπάλους τους.



Σχήμα 1α, 1β: Κατώφλια αποδοχής των δεκτών στο ultimatum game με και χωρίς δυνατότητα δημιουργίας φήμης. Πηγή: E. Fehr & U. Fischbacher (2003), "The nature of human altruism." ***Nature*** **425: 785 – 791.**

2.1.3 Τιμωρία από τρίτο πρόσωπο

Τα προηγούμενα πειράματα είχαν σαν στόχο τη μελέτη της συμπεριφοράς των ατόμων σε συνθήκες όπου παραβιάζονται κάποια συμπεριφορικά πρότυπα δίκαιης κατανομής και συνεργασίας. Παρατηρήθηκε μια τάση των ατόμων να τιμωρήσουν την παραβατική συμπεριφορά επωμιζόμενα ένα κόστος για να επιβάλουν κυρώσεις, παρόλο που εκλείπουν τα κίνητρα για κάποιο προσδοκώμενο υλικό όφελος. Αυτή η προδιάθεση των ατόμων για αλτρουιστική τιμωρία φαίνεται να αποτελεί ένα σημαντικό μηχανισμό για την επιβολή και την ενίσχυση των κανόνων κοινωνικής συμπεριφοράς.

Στις ανωτέρω περιπτώσεις τα άτομα που τιμωρούν συμμετέχουν σαν μέρη στην αλληλεπίδραση και το οικονομικό τους όφελος επηρεάζεται άμεσα από την παραβίαση του κοινωνικού προτύπου. Στην κοινωνία ωστόσο, οι αλληλεπιδράσεις αποτελούν αντικείμενο παρατήρησης και από άλλα, ΄΄τρίτα΄΄ μέρη τα οποία δε συμμετέχουν σε αυτές και δε δέχονται κάποια επιρροή στο οικονομικό τους όφελος από αυτές. Σε αυτήν την περίπτωση το κάθε μέρος παρατηρεί τη συμπεριφορά των άλλων και συχνά μια άδικη συμπεριφορά μπορεί να επιφέρει τιμωρία από ένα μέρος το οποίο δε θίγεται άμεσα από αυτήν. Προκύπτει λοιπόν το ερώτημα αν το αίσθημα διόρθωσης της αδικίας επεκτείνεται και σε αλληλεπιδράσεις στις οποίες τα άτομα δε συμμετέχουν άμεσα, αν δηλαδή αυτή η έμφυτη ροπή έχει καθολική ισχύ για όλες τις καταστάσεις στις οποίες το άτομο κρίνει ότι έχει συμβεί μια αδικία ή παραβίαση των αποδεκτών κοινωνικών κανόνων.

Στην πραγματική ζωή το φαινόμενο της επιβολής κυρώσεων από τρίτα μέρη (*third-party sanctioning*) έχει καταγραφεί σε εθνογραφικές μελέτες για την επιβολή κοινωνικών κανόνων σε κοινωνίες μικρής κλίμακας (Fessler, 2002; Hill, 2002; Bendor & Swistak, 2001). Η ύπαρξη της είναι πολύ σημαντική για την ενίσχυση των κοινωνικών κανόνων γιατί συχνά οι παραβιάσεις επηρεάζουν ένα μόνο ή λίγα μέρη και το κόστος επιβολής της τιμωρίας θα ήταν πολύ μεγάλο για ένα μεμονωμένο άτομο. Σε συνθήκες όπου απαιτείται συνεργατική προσπάθεια μάλιστα μπορούμε να θεωρήσουμε ότι, όταν η ομάδα είναι αρκετά μεγάλη, η επιρροή στο ατομικό οικονομικό όφελος του κάθε μέρους από την έλλειψη συνεργασίας ενός λαθρεπιβάτη επιμερίζεται τόσο που είναι αμελητέα. Θα μπορούσαμε δηλαδή να πούμε ότι κανείς δεν αδικείται άμεσα σε τέτοιο βαθμό που να προκαλέσει μια αντίδραση λόγω της ατομικής μείωσης

οφέλους που υφίσταται. Άρα η ύπαρξη μιας τάσης για επιβολή τιμωρίας σε κάθε παρατηρούμενη παραβίαση που θίγει ένα μέλος ή το σύνολο μιας ομάδας ή της κοινωνίας, ακόμα και όταν το άτομο δεν είναι άμεσα θιγόμενο, είναι πολύ σημαντική προκειμένου να ενισχυθεί η συνεργασία και οι κοινωνικοί κανόνες.

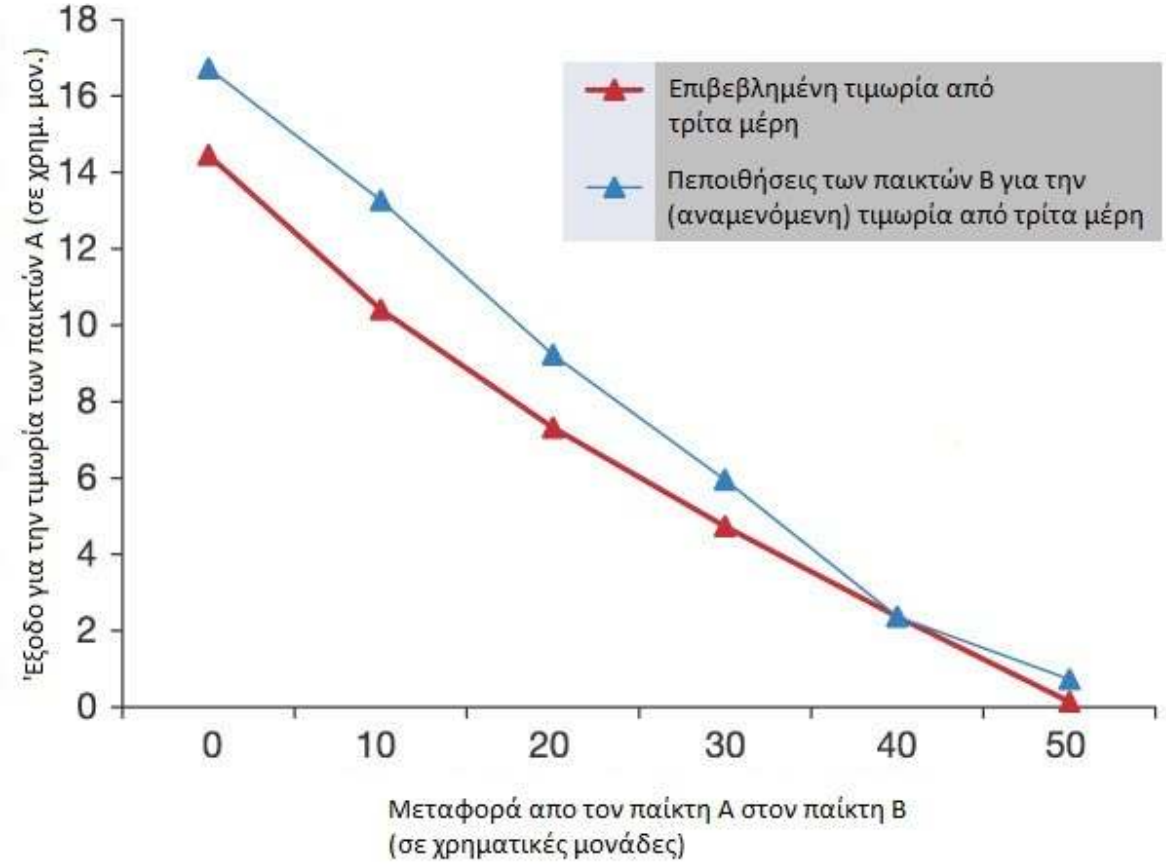
Η απομόνωση ωστόσο των παραγόντων που οδηγούν σε μια τέτοια παρατηρούμενη συμπεριφορά τιμωρίας από τρίτα πρόσωπα ή μέρη (*third-party punishment*) στην πραγματική ζωή είναι εξαιρετικά δύσκολη, όπως άλλωστε αναφέρθηκε προηγουμένως και για την τιμωρία από δεύτερα μέρη (*second-party punishment*). Και πάλι είναι στην πράξη αδύνατο να διαπιστώσουμε κατά πόσο μια κύρωση οφείλεται σε εγωιστικά κίνητρα άμεσης αμοιβαιότητας ή δημιουργίας φήμης και κατά πόσο σε προδιάθεση αλτρουιστική τιμωρίας. Το ερώτημα μπορεί να ελεγχθεί σε πειράματα τιμωρίας από τρίτο πρόσωπο που διεξάγονται σε εργαστηριακές συνθήκες.

Στο σχεδιασμό των πειραμάτων τιμωρίας από τρίτο πρόσωπο εισάγεται ένα τρίτο άτομο (τρίτο μέρος) το οποίο παρακολουθεί την ανώνυμη αλληλεπίδραση μεταξύ δύο άλλων μερών. Μεταξύ των συμμετεχόντων υπάρχει απόλυτη ανωνυμία, κανείς δηλαδή δεν έχει πληροφορία για το ιστορικό συμπεριφοράς του άλλου ατόμου και οι αλληλεπιδράσεις είναι one-shot, τα άτομα δηλαδή γνωρίζουν ότι δεν πρόκειται να ξανασυναντηθούν. Το τρίτο μέρος αφού ολοκληρωθεί η αλληλεπίδραση έχει τη δυνατότητα, επωμιζόμενο ένα κόστος, να επιβάλει τιμωρία (χρηματική ποινή) στα άλλα μέρη. Αν το άτομο λειτουργεί με αμιγώς εγωιστικά κίνητρα μεγιστοποίησης του ατομικού του οικονομικού οφέλους δεν έχει ποτέ λόγο να τιμωρήσει, εφόσον αυτό θα μειώσει το όφελός του χωρίς να του προσφέρει κανενός είδους κέρδος. Ωστόσο, αν το άτομο έχει μια ισχυρή έμφυτη τάση αντίδρασης απέναντι στην συμπεριφορά που παραβιάζει τους κανόνες, θα επιβάλει κυρώσεις παρά το κόστος.

Ένα παράδειγμα τέτοιου παιγνίου είναι ένα παιχνίδι του δικτάτορα με τρεις παίκτες. (Fehr & Fischbacher, 2004). Ο παίκτης Α, ο δικτάτορας έχει στη διάθεσή του 100 χρηματικές μονάδες, από τις οποίες μπορεί να μεταφέρει ένα οποιοδήποτε μέρος στον παίκτη Β, τον αποδέκτη. Ο παίκτης Β δεν έχει χρηματικό διαθέσιμο και είναι υποχρεωμένος να δεχτεί την προσφορά του Α, όποια και αν είναι αυτή. Ο παίκτης Γ απλά παρακολουθεί τη μεταφορά και έχει και ο ίδιος στη διάθεσή του 50 χρηματικές μονάδες. Μετά τη μεταφορά μπορεί να επιβάλει πόντους ποινής στον παίκτη Α. Για κάθε πόντο ποινής ο παίκτης Γ πρέπει να θυσιάσει 1 χρηματική μονάδα και ο παίκτης Α χάνει 3 χρηματικές μονάδες. Τα αποτελέσματα του πειράματος μας δείχνουν ότι οι

παίκτες Α δεν τιμωρούνται ποτέ αν μεταφέρουν 50 ή περισσότερες μονάδες στον παίκτη Β. Όταν μεταφέρουν κάτω από 50 μονάδες, το 55% των παικτών Γ τους τιμωρεί και η τιμωρία αυξάνεται όσο μικρότερη γίνεται η χρηματική μεταφορά του Α (Σχήμα 2). Φαίνεται λοιπόν ότι οι παίκτες Γ δε λειτουργούν εγωιστικά, αλλά όταν εντοπίζουν συμπεριφορά που παραβιάζει το πρότυπο δικαιοσύνης (μεταφορά κάτω από το μισό του διαθέσιμου ποσού), αποφασίζουν να πληρώσουν ένα αντίτιμο για να επιβάλουν ποινή στον 'άδικο' παίκτη. Οι ίδιοι δεν έχουν να περιμένουν οποιοδήποτε κέρδος από την πράξη αυτή, επομένως πρόκειται για μια πράξη αλτρουιστικής τιμωρίας.

Την ύπαρξη αυτού του μηχανισμού αντίδρασης την επιβεβαιώνει και το γεγονός ότι οι παίκτες Β σε ποσοστό 70-80% δηλώνουν σε ερωτηματολόγιο ότι αναμένουν αυτήν την τιμωρία. Αυτό αποτελεί ένδειξη ότι η ψυχολογική αυτή πραγματικότητα είναι κοινή ανάμεσα στους ανθρώπους. Όταν ο παίκτης Α δε μετέφερε τίποτα στον Β, δεχόταν κατά μέσο όρο 9 πόντους ποινής από τους παίκτες Γ και επομένως το εισόδημά του μειωνόταν κατά 3 x 9 = 27 μονάδες. Αυτό σημαίνει ότι από μια εγωιστική σκοπιά, θα ήταν συμφέρον για τον παίκτη να μην κάνει καμία μεταφορά, σε σύγκριση με το να κάνει μια μεταφορά κοντά στο 50% προκειμένου να αποφύγει το κόστος της ποινής. Ωστόσο αν υπάρχουν πολλοί παίκτες Γ που μπορούν να τιμωρήσουν ταυτόχρονα τον Α αυτό πλέον δεν ισχύει, γιατί η συνολική ποινή που δέχεται ο Α αυξάνεται. Γίνεται λοιπόν εμφανής η σημασία που έχει για τη επιβολή των κανόνων η παρουσία ατόμων με προδιάθεση να τιμωρήσουν τους παραβάτες σε μια κοινωνία.



Σχήμα 2: Αλτρουιστική τιμωρία απο τρίτα μέρη που δεν επηρεάζονται άμεσα από την παραβίαση ενός κοινωνικού κανόνα δίκαιης κατανομής. Πηγή: E. Fehr & U. Fischbacher (2004), "Third-party punishment and social norms." ***Evolution and Human Behavior*** **25: 63-87.**

Παρόμοια αποτελέσματα λαμβάνουμε και σε πειράματα διλήμματος του φυλακισμένου όπου ένα τρίτο πρόσωπο παρατηρεί ένα ζεύγος που παίζει το παιχνίδι (Fehr & Fischbacher, 2004). Το τρίτο πρόσωπο είναι εξοικειωμένο με τους όρους του παιχνιδιού, καθώς κατά τη διάρκεια του πειράματος παίζει και αυτό το παιχνίδι με ένα διαφορετικό συμπαίκτη. Το τρίτο πρόσωπο μπορεί μετά το παιχνίδι να τιμωρήσει -με κόστος για το ίδιο- τον παίκτη του άλλου ζεύγους που παρακολουθεί. Και πάλι παρατηρούμε ότι τα τρίτα μέρη επιβάλλουν ποινή όταν αντιλαμβάνονται συμπεριφορά που παραβιάζει το συνεργατικό πρότυπο, όταν δηλαδή βλέπουν κάποιον να προδίδει ενώ ο συμπαίκτης του έχει συνεργαστεί. Τα επίπεδα τιμωρίας από το τρίτο μέρος είναι πολύ υψηλά, σχεδόν το ίδιο υψηλά με την τιμωρία από το δεύτερο μέρος, δηλαδή τον άμεσα συμμετέχοντα. Το γεγονός ότι η επιβολή τιμωρίας από τρίτο μέρος συμβαίνει και σε πλαίσια όπου πρέπει να προωθηθούν συνεργατικοί κανόνες και να τιμωρηθούν οι λαθρεπιβάτες

είναι πολύ σημαντικό για την κοινωνία, γιατί πολλές φορές στην πραγματική ζωή η επίδραση του λαθρεπιβάτη είναι πολύ μικρή για το όφελος των μεμονωμένων ατόμων. Η τάση για τιμωρία τους από τα άλλα άτομα παρά το στοιχείο αυτό αυξάνει την ικανότητα της κοινωνίας για επιβολή των συνεργατικών προτύπων.

2.1.4. Παίγνια δημόσιων αγαθών

Τα παίγνια που περιγράφονται παραπάνω μας δίνουν ισχυρές ενδείξεις για την ύπαρξη μιας εσωτερικής τάσης των ατόμων να τιμωρήσουν τις άδικες συμπεριφορές σε μια αλληλεπίδραση και της επιθυμίας να εφαρμόσουν αυτή την τιμωρία αναλαμβάνοντας ένα κόστος, ακόμα και όταν δεν προσδοκούν κάποιο άμεσο ή έμμεσο όφελος. Οι αλληλεπιδράσεις είναι διμερείς και η τιμωρία που μπορεί να επιβάλει το ένα μέλος επηρεάζει άμεσα το άλλο, οδηγώντας το σε μεγαλύτερη συνεργατικότητα και συμμόρφωση με τους κοινωνικούς κανόνες δίκαιης συμπεριφοράς. Στην κοινωνία ωστόσο οι αλληλεπιδράσεις δεν περιορίζονται σε δύο μέλη αλλά μπορεί να περιλαμβάνουν μεγαλύτερο αριθμό ατόμων. Είναι λοιπόν εύλογο να προσπαθήσουμε να εντοπίσουμε ενδείξεις για την ύπαρξη συμπεριφοράς με τάσεις αλτρουιστικής τιμωρίας και σε αλληλεπιδράσεις μεταξύ πολλών ατόμων, καθώς και ποια είναι η επίδρασή της στη συνεργασία.

Τα παίγνια δημόσιων αγαθών (*public good games*) προσφέρουν τη δυνατότητα για μια τέτοια μελέτη (Ledyard, 1995). Βασικό χαρακτηριστικό των δημόσιων αγαθών είναι ότι κανείς δε μπορεί να αποκλειστεί από την κατανάλωσή τους. Τα δημόσια αγαθά μπορούν να καταναλωθούν από κάθε μέλος της ομάδας, ανεξάρτητα από τη συνεισφορά του μέλους στο αγαθό. Για αυτό το κάθε μέλος έχει κίνητρο να φερθεί ως λαθρεπιβάτης και να επωφεληθεί από τις συνεισφορές των άλλων, χωρίς ο ίδιος να συμβάλει στη δημιουργία του αγαθού. Αν θεωρήσουμε ότι υπάρχουν συμπεριφορές ισχυρής αμοιβαιότητας, θα περιμέναμε ότι κάτω από αυτές τις συνθήκες η αλτρουιστική ανταμοιβή θα έκανε τα άτομα να αυξήσουν τις ατομικές τους συνεισφορές εάν οι αναμενόμενες συνεισφορές από τα υπόλοιπα μέλη της ομάδας αυξάνονται. Τα άτομα θα ανταμείβουν τους άλλους αν περιμένουν ότι αυτοί θα ανεβάσουν το επίπεδο συνεργασίας τους. Αντίστοιχα, αν τα άτομα παρατηρούν ότι οι άλλοι ‘προδίδουν’ τη συνεργασία και υπάρχει η δυνατότητα να τους τιμωρήσουν, τότε θα το κάνουν ακόμα και αν πρέπει να επωμιστούν κόστος. Αν τα άτομα το περιμένουν αυτό, θα έχουν κίνητρο να συνεργαστούν ακόμα

και αν οι τάσεις συμπεριφοράς τους είναι εγωιστικές. Αναμένουμε δηλαδή να παρατηρήσουμε μια αύξηση των επιπέδων συνεργασίας όταν υπάρχει δυνατότητα τιμωρίας.

Οι κανόνες ενός χαρακτηριστικού παιγνίου δημόσιου αγαθού (Fehr & Gächter, 2000) έχουν ως εξής: Σε μια ομάδα με τέσσερις αλληλεπιδρώντες συμμετέχοντες, ο κάθε συμμετέχων λαμβάνει ένα χρηματικό διαθέσιμο 20 χρηματικών μονάδων. Οι συμμετέχοντες αποφασίζουν *συγχρόνως* πόσες μονάδες θα κρατήσουν για τον εαυτό τους και πόσες θα επενδύσουν στο κοινό αγαθό. Κάθε μονάδα που κρατάει το άτομο συνυπολογίζεται στην τελική ατομική του απολαβή. Κάθε μονάδα που επενδύει το άτομο στο κοινό αγαθό έχει σαν αποτέλεσμα να αυξηθεί η τελική ατομική απολαβή κάθε συμμετέχοντα της ομάδας κατά 0,4 μονάδες. Επομένως, η συνολική ατομική απόδοση για κάθε μονάδα που προστίθεται στο κοινό αγαθό είναι -1+0,4 = 0,6 μονάδες. Αυτό σημαίνει ότι, άσχετα από το πόσο συνεισφέρουν στο αγαθό οι άλλοι τρεις συμμετέχοντες, είναι πάντα καλύτερο για ένα συμμετέχοντα να κρατήσει όλες του τις μονάδες για τον εαυτό και να μη συνεισφέρει τίποτα στο αγαθό. Αν δηλαδή όλοι οι συμμετέχοντες συμπεριφέρονταν πλήρως εγωιστικά, δε θα συνεισέφεραν τίποτα στο αγαθό και θα κρατούσαν τις μονάδες τους περιμένοντας να επωφεληθούν επιπλέον από τις τυχόν συνεισφορές των άλλων. Ωστόσο, αν όλοι προδώσουν πλήρως τη συνεργασία κρατώντας όλες τις μονάδες τους, ο καθένας κερδίζει μόνο 20 χρηματικές μονάδες, ενώ αν όλοι επενδύσουν το συνολικό τους χρηματικό διαθέσιμο, ο κάθε συμμετέχων κερδίζει 0,4 x (4 x 20) = 32 χρηματικές μονάδες.

Στο πείραμα υπάρχουν δύο συνθήκες: η συνθήκη όπου δεν υπάρχει δυνατότητα τιμωρίας και η συνθήκη όπου υπάρχει δυνατότητα τιμωρίας. Και στις δύο συνθήκες η ίδια ομάδα συμμετεχόντων παίζει το παιχνίδι για δέκα περιόδους και στο τέλος κάθε περιόδου πληροφορούνται σχετικά με τις συνεισφορές των άλλων τριών μελών της ομάδας. Στη συνθήκη με τιμωρία όμως, εκτός από την απόφαση για το τι θα επενδύσουν, οι συμμετέχοντες μπορούν επίσης να να επιβάλουν πόντους ποινής στα άλλα μέλη της ομάδας στο τέλος κάθε περιόδου και αφού ενημερωθούν για τις συνεισφορές των άλλων. Το κόστος της τιμωρίας για τον «τιμωρό» είναι μια αύξουσα κυρτή συνάρτηση του συνολικού αριθμού πόντων ποινής που επέβαλλε στους άλλους. Το ανώτατο όριο πόντων ποινής που μπορούν να επιβληθούν από ένα άτομο σε ένα άλλο είναι δέκα. Έτσι, η επιβολή δέκα πόνων ποινής σε ένα άλλο μέλος κοστίζει στον τιμωρό 30 χρηματικές μονάδες, η επιβολή καθόλου πόντων δεν κοστίζει τίποτα και η επιβολή ενός ενδιάμεσου αριθμού πόντων ποινής έχει ένα ενδιάμεσο κόστος. Για κάθε πόντο ποινής που δέχεται ένας συμμετέχων, υφίσταται μια μείωση 10% στην τελική χρηματική του απολαβή. Μια

μείωση 10% αντιστοιχεί, κατά μέσο όρο, σε μια μείωση εισοδήματος μεταξύ δύο και τριών πόντων. Κατά τη διεξαγωγή του πειράματος διαβεβαιώνεται ότι τα μέλη της ομάδας δε μπορούν να μάθουν τίποτα για το ιστορικό των ατομικών επενδύσεων ή των ατομικών ποινών των άλλων συμμετεχόντων. Επομένως, οι αλληλεπίδραση είναι ανώνυμη και δεν είναι δυνατόν να κερδίσει κανείς φήμη για το ότι είναι συνεργατικός ή μη ή για το ότι είναι τιμωρός.

Εκτός από τις δύο διαφορετικές συνθήκες (με τιμωρία και χωρίς), το πείραμα διεξήχθη και κάτω από δύο διαφορετικού σχεδιασμούς: στον ένα η σύνθεση της ομάδας ήταν σταθερή, ενώ στον άλλο η σύνθεση της ομάδας άλλαζε σε καθεμία από τις δέκα περιόδους με τυχαία αναδιάταξη των τετράδων που συμμετείχαν. Όταν η σύνθεση της ομάδας είναι τυχαία σε κάθε περίοδο η πιθανότητα να συναντηθούν τα ίδια μέλη ομάδας σε μελλοντικές περιόδους είναι πολύ μικρή. Επιπλέον, ακόμη και αν κάποια από τα μέλη της ομάδας σε κάποια περίοδο έχουν ήδη συναντηθεί σε μια από τις προηγούμενες περιόδους, δεν είναι σε θέση να το αναγνωρίσουν το ένα το άλλο καθώς δεν υπάρχει πληροφορία για την ταυτότητα και το ιστορικό τους. Επομένως, ο σχεδιασμός τυχαίας σύνθεσης στην ουσία αποτελεί μια κατάσταση στην οποία ξένοι αλληλεπιδρούν ανώνυμα σε μια σειρά one-shot παιγνίων.

Τα ευρήματα αυτού του πειράματος έχουν μεγάλο ενδιαφέρον για τη θεωρία της ισχυρής αμοιβαιότητας και την αλτρουιστική τιμωρία, μιας και περιμένουμε να δούμε πολύ διαφορετικά αποτελέσματα αν τα άτομα συμπεριφέρονται πλήρως εγωιστικά σε σύγκριση με την περίπτωση όπου υπάρχουν άτομα με ισχυρές προτιμήσεις που αφορούν τους άλλους (*other-regarding preferences*) που συμπεριφέρονται ως αλτρουιστικοί τιμωροί. Στο συγκεκριμένο παίγνιο όταν μιλάμε για συνεργατική συμπεριφορά των συμμετεχόντων εννοούμε τη συνεισφορά τους στο δημόσιο αγαθό. Οι συνεισφορές τους λοιπόν στο δημόσιο αγαθό και ο τρόπος που αυτές μεταβάλλονται από περίοδο σε περίοδο, καθώς και το ύψος της τιμωρίας και τα άτομα προς τα οποία αυτή κατευθύνεται στη συνθήκη με τιμωρία, μπορούν να μας δώσουν ενδείξεις για την ύπαρξη αλτρουιστικής τιμωρίας και την επίδρασή της στη συνεργασία εντός ομάδων. Παρακάτω γίνεται μια αναλυτική παραγραφή των προβλέψεων που δημιουργούν οι προϋποθέσεις του πειράματος.

Στο σχεδιασμό τυχαίας σύνθεσης ομάδας η πρόβλεψη είναι αρκετά απλή αν υποθέσουμε ότι όλα τα άτομα είναι ορθολογικοί εγωιστές και αυτό αντανακλάται και στις προσδοκίες των μελών της ομάδας σχετικά με τη συμπεριφορά των υπολοίπων. Το σημαντικό στοιχείο αυτού του σχεδιασμού είναι ότι τα μέλη της ομάδας αλλάζουν σε κάθε περίοδο, άρα οι συμμετέχοντες

ξέρουν ότι σε κάθε επόμενη περίοδο θα παίζουν με νέους συμπαίκτες (ακόμα και αν είναι οι ίδιοι δεν υπάρχει τρόπος να το γνωρίζουν). Στη συνθήκη *χωρίς τιμωρία* για ένα εγωιστικό συμμετέχοντα δεν έχει νόημα να συνεργαστεί σε καμία περίοδο, επειδή η συνεισφορά στο δημόσιο αγαθό μειώνει το ατομική του χρηματική απολαβή. Εφόσον σε κάθε περίοδο οι συμμετέχοντες στην ουσία παίζουν ένα one-shot παίγνιο, στη συνθήκη *με τιμωρία* δε θα τιμωρήσουν ποτέ επειδή η τιμωρία έχει κόστος και δεν αποφέρει μελλοντικά οφέλη. Παρόλο που είναι πιθανόν η τιμωρία να κάνει τον τιμωρούμενο συμμετέχοντα να αυξήσει τη συνεργασία στις μελλοντικές περιόδους, για τον τιμωρό η πιθανότητα να κερδίσει από αυτό είναι πολύ μικρή επειδή και η πιθανότητα να ξανασυναντήσει τον τιμωρούμενο είναι μικρή. Επομένως, σε αυτό το σχεδιασμό, η τιμωρία άλλων συμμετεχόντων δεν έχει νόημα για ένα εγωιστικό άτομο. Αφού αφαιρείται η απειλή τιμωρίας, ούτε στη συνθήκη με τιμωρία έχουν τα εγωιστικά άτομα κάποιο κίνητρο να συνεισφέρουν. Άρα, και στις δύο συνθήκες του σχεδιασμού τυχαίας σύνθεσης δεν περιμένουμε να υπάρξει συνεργασία, δηλαδή συνεισφορές στο δημόσιο αγαθό, αν όλοι οι συμμετέχοντες είναι πλήρως εγωιστικά άτομα. Επίσης περιμένουμε να μην επιβληθούν τιμωρίες στη συνθήκη τιμωρίας.

Ωστόσο, τα πράγματα είναι διαφορετικά αν οι συμμετέχοντες που φέρονται εγωιστικά περιμένουν ότι στην ομάδα υπάρχουν κάποιοι συμμετέχοντες που συμπεριφέρονται σύμφωνα με τη θεωρία ισχυρή αμοιβαιότητας, έχουν δηλαδή την τάση να τιμωρούν τις μη συνεργατικές συμπεριφορές ακόμη και όταν αυτό έχει κόστος και δεν αποφέρει όφελος. Αν λοιπόν υπάρχουν αυτοί οι αλτρουιστικοί τιμωροί οι οποίοι θα έχουν την τάση να συνεισφέρουν στο δημόσιο αγαθό, οι συμμετέχοντες θα περιμένουν ότι η συμπεριφορά του λαθρεπιβάτη, δηλαδή η μη συνεισφορά στο δημόσιο αγαθό είναι πολύ πιθανό να τιμωρηθεί. Επομένως, στη συνθήκη με τιμωρία θα έχουν κίνητρο για να συνεργαστούν. Έτσι λοιπόν, όταν υπάρχουν αλτρουιστικοί τιμωροί, περιμένουμε να δούμε διαφορετικά επίπεδα συνεργασίας μεταξύ των συνθηκών χωρίς και με τιμωρία. Στη συνθήκη με τιμωρία περιμένουμε περισσότερη συνεργασία απ’ ό,τι στη συνθήκη χωρίς τιμωρία.

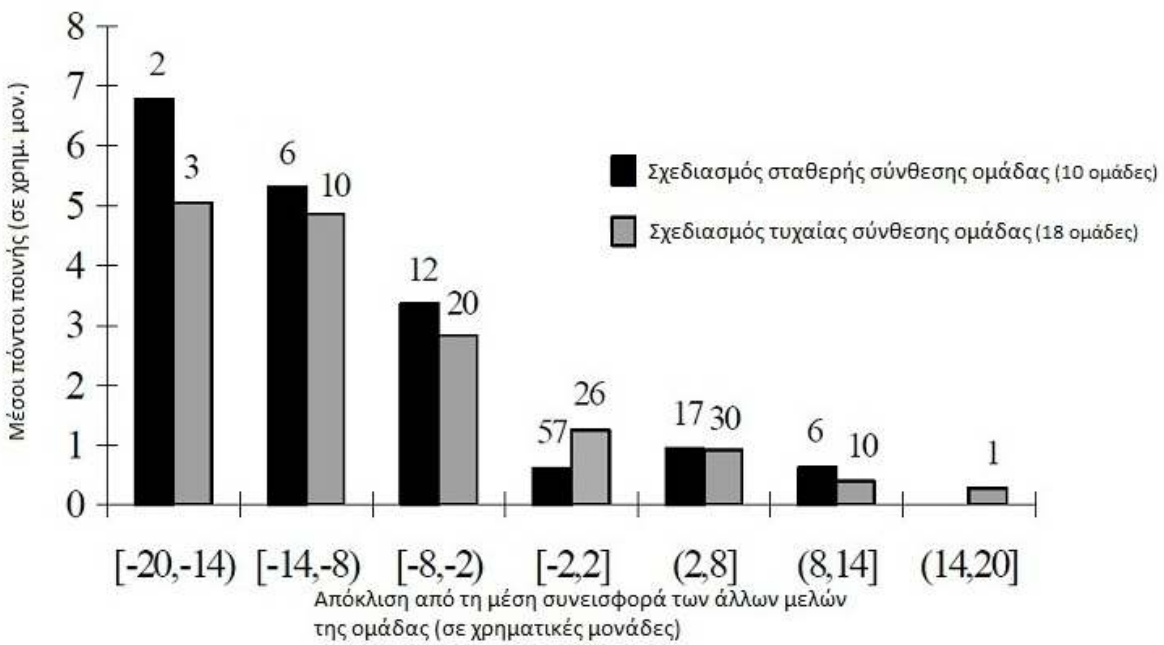
Στο σχεδιασμό σταθερής σύνθεσης ομάδας, όπου οι συμμετέχοντες παίζουν σε όλες τις περιόδους με τους ίδιους συμπαίκτες, οι προβλέψεις είναι παρόμοιες με αυτές του σχεδιασμού τυχαίας σύνθεσης, αν είναι κοινή γνώση ότι όλοι οι συμμετέχοντες είναι ορθολογικοί εγωιστικοί μεγιστοποιητές. Περιμένουμε δηλαδή να μην υπάρχει καμία συνεισφορά στο δημόσιο αγαθό τόσο στη συνθήκη χωρίς τιμωρία όσο και στη συνθήκη με τιμωρία, καθώς και καμία επιβολή

τιμωρίας στη συνθήκη με τιμωρία. Η πρόβλεψη αυτή βασίζεται σε ένα συλλογισμό οπισθοβατικής επαγωγής (*backward induction*). Η επιλογή της μη συνεισφοράς είναι πιο προφανής για τη δέκατη και τελευταία περίοδο, στην οποία ό,τι και να κάνουν οι συμμετέχοντες δεν έχει πλέον καμία επίδραση στη συμπεριφορά των συμπαικτών τους, αφού δεν υπάρχει επόμενη περίοδος. Εφόσον όλοι οι εγωιστικοί συμμετέχοντες γνωρίζουν ότι το πείραμα τελειώνει στη δέκατη περίοδο, η καλύτερη ατομική επιλογή στη συνθήκη *χωρίς τιμωρία* είναι να μη συνεισφέρουν τίποτα. Στη συνθήκη *με τιμωρία* η καλύτερη εγωιστική επιλογή στο στάδιο τιμωρίας της δέκατης περιόδου είναι να μην επιβάλουν καθόλου τιμωρία σε κανένα αφού η τιμωρία έχει κόστος και δε μπορεί πια να προσδώσει κάποιο επιπλέον όφελος. Αφού όμως οι συμμετέχοντες ως ορθολογικοί εγωιστές περιμένουν ότι κανείς δε θα τιμωρήσει, το στάδιο τιμωρίας δε μεταβάλλει τα κίνητρα στο στάδιο συνεισφοράς της δέκατης περιόδου, εφόσον στην ουσία εκλείπει η απειλή της τιμωρίας. Έτσι, και στη συνθήκη με τιμωρία, κανείς δε θα συνεισφέρει τίποτα στη δέκατη περίοδο. Εφόσον όμως οι ορθολογικοί εγωιστές περιμένουν αυτό το αποτέλεσμα για τη δέκατη περίοδο, γνωρίζουν ότι οι πράξεις τους στη στην ένατη περίοδο δεν έχουν καμία επίδραση στις αποφάσεις της δέκατης περιόδου. Επομένως, η τιμωρία στην ένατη περίοδο δεν έχει νόημα για εγωιστές συμμετέχοντες και συνεπώς η μη συνεισφορά στο αντίστοιχο στάδιο της ένατης περιόδου είναι και πάλι η καλύτερη επιλογή. Αυτό το επιχείρημα οπισθοβατικής επαγωγής μπορεί να επαναληφθεί μέχρι την πρώτη περίοδο, και έτσι η πρόβλεψη για όλες τις περιόδους της συνθήκης με τιμωρία είναι ότι δε θα υπάρχουν καθόλου συνεισφορές και καμία επιβολή τιμωρίας. Η ίδια ακριβώς λογική μπορεί να εφαρμοστεί φυσικά και για τη συνθήκη χωρίς τιμωρία, όπου επίσης προβλέπουμε να μην υπάρχει καθόλου συνεργασία αν όλοι οι συμμετέχοντες είναι εγωιστικά άτομα.

Η παρουσία ατόμων που συμπεριφέρονται σύμφωνα με την ισχυρή αμοιβαιότητα, όπως και στο σχεδιασμό τυχαίας σύνθεσης, αλλάζει τις προβλέψεις σημαντικά στο σχεδιασμό σταθερής σύνθεσης. Στη συνθήκη χωρίς τιμωρία οι συμμετέχοντες αυτοί θα παρουσιάζουν τάση να συνεισφέρουν, τουλάχιστον στην αρχή, και θα έχουν και μια τάση να συνεχίσουν να συνεισφέρουν σαν επιβράβευση στα άλλα μέλη της ομάδας που έκαναν το ίδιο. Στη συνθήκη με τιμωρία περιμένουμε ακόμα μεγαλύτερη συνεργασία. Ο λόγος είναι ότι αν ένας συμμετέχων τιμωρείται για συμπεριφορά λαθρεπιβάτη σε κάποια περίοδο πριν τη δέκατη, γνωρίζει ότι ο τιμωρός θα είναι σίγουρα μέλος της ομάδας και στην επόμενη περίοδο. Επομένως, το τιμωρούμενο μέλος, ακόμα και αν επιδεικνύει μια πλήρως εγωιστική συμπεριφορά, έχει πολύ

ισχυρό κίνητρο να συνεισφέρει στις επόμενες περιόδους λόγω της απειλής της τιμωρίας. Στο σχεδιασμό σταθερής σύνθεσης επομένως τα άτομα έχουν υψηλότερα κίνητρα συνεργασίας από ό,τι στο σχεδιασμό τυχαίας σύνθεσης και αυτό περιμένουμε να αντικατοπτριστεί στα αποτελέσματα.

Αυτό που παρατηρούμε τελικά είναι ότι και στους δύο σχεδιασμούς υπάρχουν κάποια επίπεδα συνεργασίας σε όλες τις συνθήκες και ότι οι συμμετέχοντες τιμωρούν πολύ συχνά σε όλες τις περιόδους και των δύο σχεδιασμών. Στο σχήμα 3 παρουσιάζεται η μέση τιμωρία που επιβλήθηκε σε έναν παίκτη σα συνάρτηση της απόκλισης της συνεισφοράς του τιμωρούμενου παίκτη από τη μέση συνεισφορά των άλλων μελών της ομάδας. Είναι αξιοσημείωτο ότι το ύψος της τιμωρίας στο σχεδιασμό τυχαίας σύνθεσης είναι σχεδόν το ίδιο με το ύψος της στο σχεδιασμό σταθερής σύνθεσης. Επίσης, η τιμωρία κυρίως επιβάλλεται από παίκτες που συνεισφέρουν σε παίκτες που δε συνεισφέρουν στο δημόσιο αγαθό. Υπάρχει μια θετική συσχέτιση ανάμεσα στην απόκλιση της συνεισφοράς ενός συμμετέχοντα από τη συνεισφορά των άλλων μελών της ομάδας και στο ύψος της τιμωρίας που δέχεται. Υπάρχουν λοιπόν ισχυρές ενδείξεις ότι οι συμμετέχοντες έχουν την τάση να επιβάλλουν αλτρουιστική τιμωρία σε μέλη της ομάδας που παραβιάζουν τον συνεργατικό κανόνα και ότι η τιμωρία αυτή εξαρτάται από το αντιλαμβανόμενο μέγεθος της παραβίασης. Είναι χαρακτηριστικό ότι τιμωρίες επιβάλλονται ακόμα και στη δέκατη περίοδο, αντίθετα από οτιδήποτε θα μπορούσε να προβλέψει η θεωρία της ορθολογικής μεγιστοποίησης.

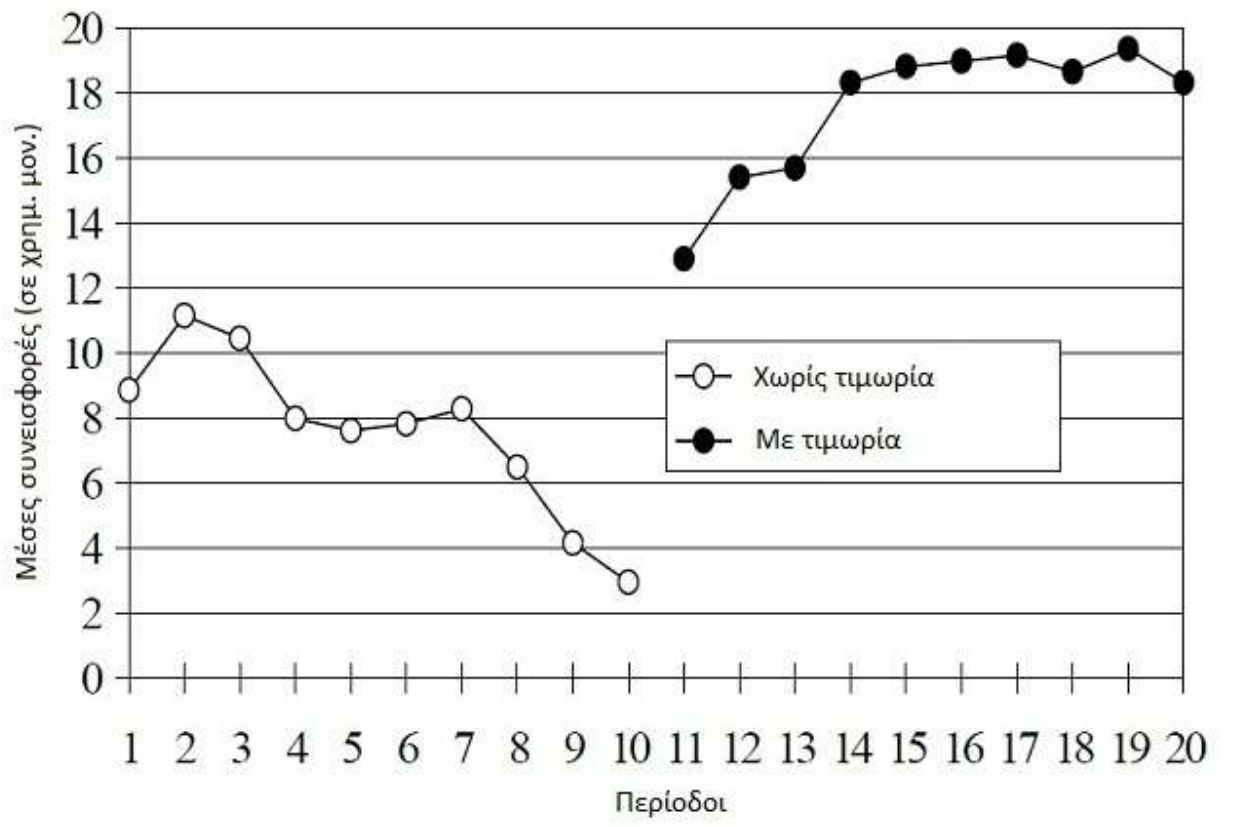


Σημ.: Οι αριθμοί πάνω από τις στήλες αποτελούν τις σχετικές συχνότητες (%) των παρατηρήσεων που αντιστοιχούν σε κάθε στήλη, των ατόμων δηλαδή που οι συνεισφορές τους είχαν τις αντίστοιχες αποκλίσεις.

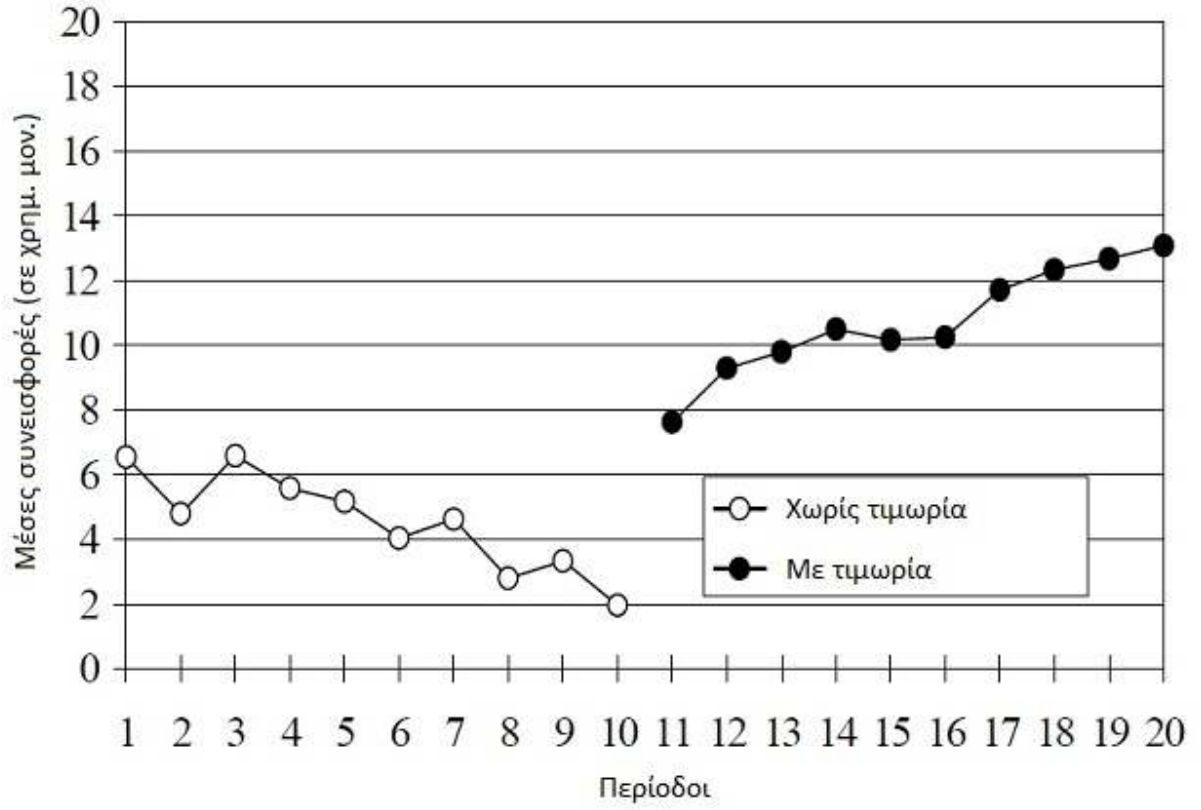
Σχήμα 3: Πόντοι ποινής που δέχτηκαν τα άτομα σε συνάρτηση με την απόκλιση της συνεισφοράς τους από τη μέση συνεισφορά των άλλων μελών της ομάδας. Πηγή: E. Fehr & S. Gächter (2000), "Cooperation and Punishment in Public Goods Experiments." ***American Economic Review*** **90: 980-994.**

Η επιβολή τιμωρίας φαίνεται ότι επηρεάζει πολύ τα επίπεδα συνεργασίας και στους δύο σχεδιασμούς. Στα σχήματα 4 και 5 φαίνεται η εξέλιξη των μέσων συνεισφορών των συμμετεχόντων από περίοδο σε περίοδο στο σχεδιασμό σταθερής και τυχαίας σύνθεσης αντίστοιχα. Το χαρακτηριστικό που παρατηρείται πολύ έντονα και στους δύο σχεδιασμούς είναι ότι η συνεργασία καταρρέει στη συνθήκη χωρίς τιμωρία. Στην αρχή έχουμε κάποια σχετικά υψηλά επίπεδα συνεργασίας, τα οποία στη συνέχεια όμως μειώνονται, και στην δέκατη περίοδο οι περισσότεροι συμμετέχοντες δε συνεισφέρουν τίποτα και κάποιοι συνεισφέρουν πολύ λίγα. Φαίνεται δηλαδή ότι η έλλειψη δυνατότητας για τιμωρία παρασύρει σε εγωιστικές συμπεριφορές και τα άτομα που αρχικά δείχνουν τάση για συμπεριφορά ισχυρής αμοιβαιότητας. Η απόσυρση της συνεισφοράς είναι ο μόνος τρόπος με τον οποίο τα άτομα αυτά μπορούν έστω έμμεσα να τιμωρήσουν τους μη συνεισφέροντες εγωιστές. Ωστόσο, όταν υπάρχει η δυνατότητα για στοχευμένη τιμωρία έναντι των λαθρεπιβατών τα αποτελέσματα αντιστρέφονται. Και στους δύο σχεδιασμούς οι αλτρουιστικοί τιμωροί επιβάλλουν ποινές στους λαθρεπιβάτες, κάτι που φαίνεται

να αυξάνει τα επίπεδα συνεργασίας. Στο σχεδιασμό σταθερής σύνθεσης μάλιστα τα επίπεδα συνεργασίας αυξάνονται κατά πολύ και στην τελευταία περίοδο σχεδόν όλοι οι συμμετέχοντες συνεισφέρουν ένα σημαντικό μέρος του διαθεσίμου τους.



Σχήμα 4: Μέση συνεισφορά κατά την εξέλιξη των περιόδων σε παίγνιο δημόσιου αγαθού με σταθερή σύνθεση ομάδας (10 ομάδες). Πηγή: E. Fehr & S. Gächter (2000), "Cooperation and Punishment in Public Goods Experiments." ***American Economic Review*** **90: 980-994.**



Σχήμα 5: Μέση συνεισφορά κατά την εξέλιξη των περιόδων σε παίγνιο δημόσιου αγαθού με τυχαία σύνθεση ομάδας (18 ομάδες). Πηγή: Fehr, E. & Gächter, S. (2000), "Cooperation and Punishment in Public Goods Experiments", ***American Economic Review*** **90: 980-994.**

Οι υψηλοί βαθμοί τιμωρίας και συνεργασίας που παρατηρούνται στις συνθήκες τιμωρίας των δύο σχεδιασμών δε μπορούν να εξηγηθούν με βάση τις θεωρίες της άμεσης και έμμεσης αμοιβαιότητας, ενώ είναι συμβατές με τη θεωρία της ισχυρής αμοιβαιότητας. Στο πείραμα αυτό αποτυπώνεται χαρακτηριστικά η τάση των ατόμων να επιβάλλουν αλτρουιστική τιμωρία όταν οι κανόνες και τα πρότυπα κοινωνικής συμπεριφοράς, εν προκειμένω της συνεργατικής συμπεριφοράς, παραβιάζονται. Η τιμωρία αυτή δε βασίζεται σε εγωιστικά κίνητρα μεγιστοποίησης της χρησιμότητας, εφόσον έχει κόστος και δεν προσφέρει οικονομικό όφελος στον τιμωρό. Η μεγάλη διαφορά στη μεταβολή των επιπέδων συνεργασίας μεταξύ των συνθηκών χωρίς και με τιμωρία δείχνει επίσης ότι η απειλή τιμωρίας μπορεί να αλλάξει τη συμπεριφορά των εγωιστών και να τους αναγκάσει να συνεργαστούν. Αντίθετα, όταν αυτή η απειλή απουσιάζει, οι εγωιστές παρασύρουν τα συνεργατικά άτομα να εγκαταλείψουν τη

συνεργασία. Η ύπαρξη επομένως ατόμων πρόθυμων να επιβάλλουν αλτρουιστική τιμωρία μπορεί να αποτελέσει ένα πολύ κρίσιμο στοιχείο για την εδραίωση και την τήρηση συνεργατικών προτύπων.

2.2 Κοινωνικές προτιμήσεις στους ανθρώπους

2.2.1 Κοινωνικές προτιμήσεις

Οι πειραματικές ενδείξεις που περιγράφηκαν παραπάνω σχετικά με την ύπαρξη της αλτρουιστικής τιμωρίας οδηγούν στο συμπέρασμα ότι τα άτομα στο εγγύς επίπεδο έχουν κίνητρα συμπεριφοράς που δεν περιορίζονται στο στενό οικονομικό και ατομικό τους συμφέρον. Φαίνεται δηλαδή ότι οι υποκειμενικές τους εκτιμήσεις των οικονομικών απολαβών δεν ταυτίζονται με την καθαρή οικονομική απολαβή, αλλά περιλαμβάνουν και άλλα στοιχεία όπως η αποστροφή προς την ανισότητα (*inequity aversion*) (Fehr & Schmidt, 1999, Dawes et al., 2007) και η αμοιβαία δικαιοσύνη (*reciprocal fairness*) (Rabin, 1993; Levine, 1998). Στην εκτίμηση τους δηλαδή περιλαμβάνεται και το ποια είναι η οικονομική απολαβή των άλλων σε μια αλληλεπίδραση, καθώς και οι προθέσεις που οι άλλοι δείχνουν με τη συμπεριφορά και τις επιλογές τους. Στην οικονομική βιβλιογραφία το γεγονός αυτό έχει αποτελέσει βάση για μια κατηγορία θεωριών που ονομάζονται θεωρίες κοινωνικών προτιμήσεων (*social preferences theories*) (Dufwenberg & Kirchsteiger, 1999; Falk & Fischbacher, 2006; Bolton & Ockenfels, 2000; Charness & Rabin, 2000).

Από θεωρητικής σκοπιάς αυτές οι προτιμήσεις δεν είναι διαφορετικές από τις προτιμήσεις π.χ. για τροφή ή για το παρόν σε σχέση με το μέλλον. Η ύπαρξη των προτιμήσεων αυτών μπορεί να εξηγήσει τις συμπεριφορές ισχυρής αμοιβαιότητας που παρατηρούνται στα πειράματα και να αποτελέσει το κίνητρο για αλτρουιστική τιμωρία ή ανταμοιβή.

Αν λοιπόν στους ανθρώπους αυτά τα κίνητρα συνυπάρχουν με τα εγωιστικά κίνητρα μεγιστοποίησης του οικονομικού οφέλους, τότε οι επιλογές τους μπορούν να περιγραφούν και να προβλεφθούν με συναρτήσεις χρησιμότητας που θα ενσωματώνουν σαν παράγοντες και τις προτιμήσεις αυτές. Φαίνεται δηλαδή ότι οι άνθρωποι ωθούνται και από τις δύο κατηγορίες κινήτρων και το ποια από τις δύο θα υπερισχύσει της άλλης και σε ποιο βαθμό εξαρτάται πιθανόν και από τις προσδοκίες τους για το ποιες θα είναι οι προτιμήσεις και οι αποφάσεις των άλλων. Το γεγονός αυτό μπορεί να επηρεάσει σε σημαντικό βαθμό την εκδήλωση ή μη

συνεργατικών συμπεριφορών. Σε ένα δίλημμα φυλακισμένου π.χ. τα άτομα μπορεί να προτιμήσουν τη συνεργασία από την προδοσία αν εκτιμούν ότι και ο συμπαίκτης τους θα κάνει το ίδιο, παρόλο που το στενό οικονομικό τους συμφέρον θα τους επέβαλλε να προδώσουν ανεξάρτητα από τι θα κάνει ο συμπαίκτης. Αυτή η υπόθεση στηρίζεται από πειραματικά ευρήματα (Hayashi et al., 1999; Kiyonari et al., 2000) και αποτελεί ένδειξη ότι τα κίνητρα της ανθρώπινης συμπεριφοράς δεν είναι τόσο στενά όσο υποθέτουν οι κλασικές θεωρίες μεγιστοποίησης της χρησιμότητας.

Ο συνδυασμός των εγωιστικών και κοινωνικών προτιμήσεων στα άτομα υποδεικνύει επίσης ότι η συμπεριφορές αλτρουιστικής τιμωρίας και ανταμοιβής εξαρτώνται από τις συνθήκες τις αλληλεπίδρασης και ότι η εκδήλωσή τους θα περιορίζεται από το κόστος που έχουν. Κάθε άτομο δηλαδή συνυπολογίζει το κόστος της αλτρουιστικής πράξης και πέρα από κάποιο όριο δεν είναι διατεθειμένο να την εκδηλώσει.

2.2.2 Πειραματικά ευρήματα

Οι ενδείξεις των πειραμάτων που περιγράφηκαν συνάδουν με τις παραπάνω παρατηρήσεις. Τα άτομα που συμπεριφέρονται σύμφωνα με τη θεωρία της ισχυρής αμοιβαιότητας ανταμείβουν και τιμωρούν σε ανώνυμες one-shot αλληλεπιδράσεις. Αυτές οι ανταμοιβές και η τιμωρία αυξάνονται όταν οι αλληλεπιδράσεις είναι επαναλαμβανόμενες ή όταν υπάρχουν δυνατότητες δημιουργίας φήμης, κάτι που μας δείχνει τη συνύπαρξη αλτρουιστικών και εγωιστικών κινήτρων στο ίδιο άτομο. Τα αλτρουιστικά κίνητρα το ωθούν να συνεργαστεί και να τιμωρήσει στις one-shot αλληλεπιδράσεις και τα εγωιστικά του κίνητρα το ωθούν να αυξήσει τις ανταμοιβές και την τιμωρία σε επαναλαμβανόμενες αλληλεπιδράσεις ή σε συνθήκες φήμης.

Αυτό φαίνεται καθαρά στο πείραμα του ultimatum με δυνατότητα δημιουργίας φήμης που περιγράφηκε παραπάνω (Fehr & Fischbacher, 2003), καθώς και στο σχεδιασμό σταθερής σύνθεσης του πειράματος δημοσίου αγαθού (Fehr & Gächter, 2000). Στο μεν πρώτο, όταν υπήρχε η δυνατότητα δημιουργίας φήμης, παρατηρήθηκε μια αύξηση του κατωφλιού απόρριψης προτάσεων, της έμμεσης δηλαδή τιμωρίας που μπορούσαν να επιβάλουν οι συμμετέχοντες. Στο δεύτερο παρατηρήθηκε ότι όταν οι αλληλεπιδράσεις με την ίδια ομάδα ήταν επαναλαμβανόμενες (σχεδιασμός σταθερής σύνθεσης), τα άτομα επέλεγαν να συνεισφέρουν περισσότερα ως ανταμοιβή στα άλλα συνεργατικά άτομα σε σχέση με το σχεδιασμό τυχαίας

σύνθεσης. Στη συνθήκη τιμωρίας μάλιστα, όπου τα άτομα μπορούν να περιμένουν υψηλότερες συνεισφορές από τους συμπαίκτες τους λόγω της απειλής της ποινής, οι συνεισφορές αυξάνονται πολύ περισσότερο. Αυτό δείχνει ότι όσο υψηλότερες είναι οι προσδοκίες σχετικά με τις συνεισφορές των άλλων, τα άτομα τείνουν να αυξάνουν και τις δικές τους συνεισφορές (Dawes, 1980; Messick & Brewer, 1983; Fischbacher et al, 2001).

Σε ό,τι αφορά το κόστος της αλτρουιστικής πράξης, αναμένουμε να παρατηρήσουμε μείωση των αλτρουιστικών πράξεων όσο το κόστος τους αυξάνεται. Το κόστος της δηλαδή θα πρέπει να κινείται μέσα σε κάποια λογικά πλαίσια προκειμένου το άτομο να την επιλέξει. Όταν το κόστος μεγαλώνει, τα άτομα πρέπει να μειώσουν περισσότερο τις ατομικές τους απολαβές για να ικανοποιήσουν τα αλτρουιστικά-κοινωνικά τους κίνητρα, επομένως για ένα δεδομένο συνδυασμό αλτρουιστικών και εγωιστικών κινήτρων τα άτομα θα επιδεικνύουν λιγότερο αλτρουιστική συμπεριφορά. Τα στοιχεία από παίγνια δικτάτορα και παίγνια δημόσιων αγαθών επιβεβαιώνουν αυτή την πρόβλεψη. Αν το ατομικό εισόδημα που πρέπει να επενδυθεί για να παραχθεί μια μονάδα του δημόσιου αγαθού αυξηθεί, οι συμμετέχοντες επενδύουν λιγότερο στο δημόσιο αγαθό (Ledyard, 1995; Isaac & Walker, 1988). Παρομοίως, αν το κόστος μεταφοράς μιας χρηματικής μονάδας στον αποδέκτη στο παίγνιο του δικτάτορα αυξηθεί, οι δικτάτορες δίνουν λιγότερα χρήματα στους αποδέκτες (Andreoni & Miller, 2002). Τα στοιχεία αυτά μας δείχνουν ότι για την ανάληψη της αλτρουιστικής τιμωρίας γίνεται ένας σαφής γνωσιακός συνυπολογισμός του οικονομικού κόστους της για τον τιμωρό. Στον υπολογισμό μάλιστα φαίνεται ότι συμπροσμετράται και η αναλογία του κόστους αυτού με τις επιπτώσεις στις υλικές απολαβές του τιμωρούμενου (Egas & Riedl, 2008). Για να επιλέξουν τα άτομα να επιβάλουν αλτρουιστική τιμωρία θα πρέπει να υπάρχει μια αναλογία κόστους προς επιπτώσεις τουλάχιστον 1:3. Για μικρότερες αναλογίες τα άτομα δεν είναι διατεθειμένα να επιβάλουν τιμωρία και για το λόγο αυτό το μικρό ποσοστό της τιμωρίας που επιβάλλεται δεν είναι σε θέση να διατηρήσει τη συνεργασία.

2.3 Νευροβιολογικές ενδείξεις

Η αλτρουιστική τιμωρία λοιπόν φαίνεται να αποτελεί μια χαρακτηριστική εκδήλωση των κοινωνικών προτιμήσεων των ανθρώπων, οι οποίες σε συνδυασμό με τις εγωιστικές τους

προτιμήσεις αποτελούν τα κίνητρα της συμπεριφοράς τους. Οι κοινωνικές προτιμήσεις των ανθρώπων αφορούν στοιχεία όπως η δικαιοσύνη και τα πρότυπα συνεργασίας, καθώς και ένα σύνολο κανόνων που αποτελούν τα πρότυπα της κοινωνικής συμπεριφοράς. Όταν αυτά τα πρότυπα παραβιάζονται όπως είδαμε, προκαλούν στους ανθρώπους την τάση να τιμωρήσουν τους παραβάτες, ακόμα και όταν αυτό έχει υλικό-οικονομικό κόστος χωρίς να διαφαίνεται κάποιο αντίστοιχο όφελος. Το συμπέρασμα που άμεσα προκύπτει είναι ότι στο εγγύς επίπεδο, το επίπεδο των πεποιθήσεων και των αξιών, θα πρέπει να λειτουργεί κάποιος γνωσιακός μηχανισμός ο οποίος βρίσκεται πίσω από αυτό το κίνητρο για αλτρουιστική τιμωρία. Ο συνήθης μηχανισμός πίσω από τα κίνητρα στο εγγύς επίπεδο είναι ότι οι άνθρωποι λαμβάνουν ικανοποίηση από μια πράξη. Μια εύλογη επομένως υπόθεση είναι ότι οι άνθρωποι προβαίνουν σε πράξεις αλτρουιστικής τιμωρίας επειδή η παράβαση των κοινωνικών κανόνων τους προκαλεί αρνητικά συναισθήματα (Sanfey et al, 2003) και η τιμωρία του παραβάτη τους γεννά ένα συναίσθημα ανακούφισης ή ικανοποίησης.

Η υπόθεση αυτή ελέγχθηκε σε ένα πείραμα με χρήση νευροαπεικονιστικών μεθόδων (de Quervain et al., 2004). Στο πείραμα αυτό οι συμμετέχοντες έπαιξαν ένα παίγνιο εμπιστοσύνης, κατά τη διάρκεια του οποίου λήφθηκαν απεικονίσεις των νευρωνικών ενεργοποιήσεων του εγκεφάλου τους με τη μέθοδο Τομογραφίας Εκπομπής Ποζιτρονίων (Positron Emission Tomography, PET). Το παίγνιο περιλάμβανε μια διαδικασία χρηματικής ανταλλαγής μεταξύ δύο παικτών, στην οποία ο ένας παίκτης (Β) έκανε μια επιλογή είτε να προδώσει είτε να ανταποδώσει την εμπιστοσύνη του άλλου παίκτη (Α). Στη συνέχεια ο παίκτης Α είχε τη δυνατότητα να επιβάλει χρηματική τιμωρία στον Β. Κατά τη διάρκεια του σταδίου κατά το οποία ο παίκτης Α αποφάσιζε για αυτήν του την ενέργεια λαμβάνονταν οι απεικονίσεις της τομογραφίας. Η υπόθεση ήταν ότι εάν πράγματι η επιβολή αλτρουιστικής τιμωρίας προκαλεί αισθήματα ικανοποίησης στα άτομα, τότε η εκδήλωσή της θα πρέπει να σχετίζεται με την ενεργοποίηση εγκεφαλικών περιοχών που συνδέονται με την επεξεργασία ανταμοιβών. Πολλές νευροαπεικονιστικές μελέτες (Knutson et al., 2000; Delgado et al., 2004) έχουν δείξει ότι το ραβδωτό σώμα είναι μια εγκεφαλική δομή που αποτελεί κρίσιμο κομμάτι του νευρωνικού κυκλώματος που σχετίζεται με την ανταμοιβή. Επιπλέον, αν η αλτρουιστική τιμωρία προκύπτει επειδή ο τιμωρός προσδοκά να λάβει ικανοποίηση μέσω αυτής, θα πρέπει να παρατηρείται ενεργοποίηση κυρίως σε εκείνες τις περιοχές του κυκλώματος που σχετίζονται με τη συμπεριφορά που κατευθύνεται από στόχους. Μια τέτοια βασική περιοχή φαίνεται να είναι το

ραχιαίο τμήμα του ραβδωτού σώματος (Schultz & Romo, 1988; O'Doherty et al., 2004). Αν μάλιστα η αλτρουιστική τιμωρία είναι μια πράξη που προκειμένου να γίνει προηγείται ένας γνωσιακός συνυπολογισμός του οικονομικού κόστους της έναντι του συναισθηματικού οφέλους που θα επιφέρει, τότε περιμένουμε να δούμε εντονότερη ενεργοποίηση και περιοχών που συνδέονται με την απαρτίωση διαφορετικών γνωστικών λειτουργιών, όπως ο προμετωπιαίος φλοιός (Bechara et al., 2000; Ramnani & Owen, 2004).

Προκειμένου να ελεγχθούν οι υποθέσεις σχεδιάστηκαν τέσσερις διαφορετικές συνθήκες για τη λήψη διαφορικών ενεργοποιήσεων. Στην πρώτη συνθήκη (Α) η επιβολή τιμωρίας είχε κόστος σε χρηματικές μονάδες για τον τιμωρό και για τον τιμωρούμενο. Στη δεύτερη συνθήκη (Β) η τιμωρία δεν είχε κόστος για τον τιμωρό, η επιβολή της δηλαδή ήταν 'δωρεάν', ενώ είχε κανονικά κόστος για τον τιμωρούμενο. Στην τρίτη συνθήκη (Γ) η επιβολή της τιμωρίας ήταν συμβολική, δεν επέφερε δηλαδή κάποιο κόστος στον τιμωρούμενο και δεν είχε κόστος ούτε για τον τιμωρό. Στην τέταρτη συνθήκη (Δ) η επιβολή της τιμωρίας είχε κόστος για τιμωρό και τιμωρούμενο, όμως στη συνθήκη αυτή ο τιμωρός ενημερωνόταν ότι η απόφαση προδοσίας ή ανταπόδοσης της εμπιστοσύνης είχε ληφθεί τυχαία από έναν υπολογιστή, δε μπορούσε δηλαδή να αποδοθεί πρόθεση στον άλλον παίκτη.

Μετρώντας τη διαφορική ενεργοποίηση των εγκεφαλικών περιοχών μεταξύ των συνθηκών Β και Γ παρατηρήθηκε εντονότερη ενεργοποίηση του κερκοφόρου πυρήνα (τμήμα του ραχιαίου ραβδωτού σώματος) στη συνθήκη Β. Το στοιχείο αυτό αποτελεί ένδειξη ότι η επιβολή τιμωρίας πράγματι προξενεί συναισθήματα ικανοποίησης και θετικής ανταμοιβής, εφόσον η μόνη διαφορά μεταξύ των δύο συνθηκών είναι η δυνατότητα που έχει ο συμμετέχων να επιβάλει αποτελεσματική τιμωρία. Η ενεργοποίηση επομένως του κερκοφόρου πυρήνα που επεξεργάζεται ανταμοιβές πρέπει να οφείλεται στο γεγονός αυτό. Αν λοιπόν ένα άτομο παίρνει συναισθηματική ανταμοιβή από την τιμωρία ενός παραβάτη, τότε θα είναι πρόθυμο να υποστεί και κάποιο κόστος προκειμένου να την επιβάλει. Πράγματι, τα αποτελέσματα του πειράματος δείχνουν ότι τα άτομα που παρουσιάζουν την εντονότερη διαφορική ενεργοποίηση του κερκοφόρου μεταξύ των συνθηκών Β και Γ είναι αυτά που επιβάλλουν τις μεγαλύτερες ποινές στη συνθήκη Α όπου η τιμωρία έχει κόστος για τον τιμωρό. Το στοιχείο αυτό αποτελεί βάσιμη ένδειξη ότι τα άτομα προσδοκούν κάποιο συναισθηματικό όφελος από την επιβολή τιμωρίας και το ύψος αυτού του οφέλους συσχετίζεται θετικά με το ύψος της αλτρουιστικής τιμωρίας που είναι διατεθειμένα να επιβάλουν.

Στη συνθήκη Δ, η απόφαση για προδοσία λαμβάνεται από έναν υπολογιστή, και επομένως το άτομο την αντιλαμβάνεται ως ένα τυχαίο γεγονός και όχι ως μια άδικη πράξη. Παλαιότερη έρευνα (Rilling, 2002) με νευροαπεικονιστικές μεθόδους έχει δείξει τη σημασία της πρόθεσης που δείχνει ο παραβάτης των κανόνων για την πρόκληση επιθυμίας για επιβολή τιμωρίας. Στην έρευνα αυτή βρέθηκε ότι όταν οι συμμετέχοντες επιτυγχάνουν αμοιβαία συνεργασία με ένα άλλο ανθρώπο σε ένα δίλημμα του φυλακισμένου, το κύκλωμα ανταμοιβής του εγκεφάλου ενεργοποιείται περισσότερο σε σχέση με μια συνθήκη όπου οι συμμετέχοντες επιτυγχάνουν αμοιβαία συνεργασία με έναν υπολογιστή. Επιπλέον βρέθηκε ότι υπάρχει μια αρνητική απόκριση του κυκλώματος όταν ο ένας συμμετέχων συνεργάζεται ενώ ο άλλος προδίδει.

Έτσι λοιπόν, στη συνθήκη Δ οι συμμετέχοντες δε νιώθουν την επιθυμία να τιμωρήσουν και οι περισσότεροι δεν επιβάλλουν καθόλου βαθμούς ποινής. Στη συνθήκη αυτή λοιπόν η τιμωρία δε θα πρέπει να προσφέρει κάποια ικανοποίηση. Πράγματι, στη σύγκριση των συνθηκών Α και Β με τη συνθήκη αυτή παρατηρήθηκε εντονότερη ενεργοποίηση του κερκοφόρου πυρήνα καθώς και του θαλάμου, μιας περιοχής που επίσης σχετίζεται στενά με την επεξεργασία χρηματικών ανταμοιβών.

Τέλος, η ενεργοποίηση του προμετωπιαίου φλοιού ήταν υψηλότερη στη συνθήκη Α από ό,τι στη συνθήκη Β, κάτι που αποτελεί ένδειξη ότι για τη λήψη της απόφασης για επιβολή τιμωρίας απαιτείται κάποιο είδος απαρτίωσης μεταξύ γνωστικών λειτουργιών. Αυτό στηρίζει την υπόθεση ότι η επιβολή αλτρουιστικής τιμωρίας ακολουθεί μια εσωτερική στάθμιση μεταξύ των συναισθηματικού οφέλους από την ικανοποίηση που νιώθει κάποιος όταν τιμωρεί τον παραβάτη των κοινωνικών κανόνων και του οικονομικού κόστους που πρέπει να υποστεί προκειμένου να το κάνει.

Προς την ίδια κατεύθυνση δείχνουν και τα ευρήματα μιας ακόμη νευροαπεικονιστικής μελέτης, όπου χρησιμοποιήθηκε η μέθοδος της Λειτουργικής Μαγνητικής Τομογραφίας (Functional Magnetic Resonance Imaging, fMRI) (Sanfey et al, 2003). Στην έρευνα αυτή λήφθηκαν απεικονίσεις των νευρωνικών ενεργοποιήσεων σε παίκτες που συμμετείχαν σε ένα ultimatum game. Τα αποτελέσματα έδειξαν πως οι άδικες προσφορές προκαλούσαν δραστηριότητα σε εγκεφαλικές περιοχές που σχετίζονταν τόσο με το συναίσθημα (πρόσθια νήσος) όσο και με συνειδητές γνωσιακές λειτουργίες (οπισθιοπλάγιος προμετωπιαίος φλοιός). Όταν μάλιστα υπήρχε απόρριψη των άδικων συμπεριφορών η διαφορική ενεργοποίηση των περιοχών που σχετίζονται με την επεξεργασία συναισθημάτων ήταν πολύ πιο αυξημένη. Πιο

αυξημένη ήταν επίσης η ενεργοποίηση των περιοχών αυτών όταν τα άτομα δέχονταν την άδικη προσφορά από έναν υπολογιστή, σε σχέση με όταν τη δέχονταν από έναν άνθρωπο. Η απόρριψη των άδικων προσφορών ήταν πολύ συχνότερη όταν ο συμπαίκτης ήταν άνθρωπος. Φαίνεται λοιπόν πως όταν τα άτομα νιώθουν ότι αδικούνται βιώνουν αρνητικά συναισθήματα και η αντίδρασή τους εξαρτάται από ένα συνδυασμό συναισθηματικών και γνωσιακών λειτουργιών. Τα συναισθήματα αυτά δείχνουν να συνδέονται και με την ενσυναίσθηση (*empathy*), την ικανότητα δηλαδή αναπαράστασης και συμμετοχής στα συναισθήματα που νιώθουν οι άλλοι. Η επεξεργασία της ενσυναίσθησης σχετίζεται με τις περιοχές επεξεργασίας συναισθημάτων που παρουσίαζαν ενεργοποίηση σε αυτή τη μελέτη (πρόσθια νήσος) και η διαφορά ανάμεσα στις συνθήκες που ο συμπαίκτης ήταν υπολογιστής ή άνθρωπος υποδηλώνουν ότι παίζει ρόλο στην εκτίμηση καταστάσεων που σχετίζονται με συνεργατικά πρότυπα (Singer, 2009).

Συνεπώς, παρατηρούμε ότι οι νευροβιολογικές ενδείξεις από πειράματα με νευροαπεικονιστικές μελέτες μας οδηγούν στο συμπέρασμα ότι πίσω από την αλτρουιστική τιμωρία κρύβονται κίνητρα συναισθηματικής ικανοποίησης του ατόμου που την επιβάλλει. Το όφελος αυτό δεν είναι οικονομικό, άρα η θεωρία της εγωιστικής μεγιστοποίησης της χρησιμότητας εξακολουθεί να μην καλύπτει τα ευρήματα των πειραματικών μελετών σχετικά με τα κίνητρα της ανθρώπινης συμπεριφοράς. Εκτός δηλαδή από τους γνωσιακούς μηχανισμούς στάθμισης κόστους και οφελών, η επιβολή τιμωριών επηρεάζεται και από συναισθηματικούς μηχανισμούς που ενεργοποιούνται μέσω του κυκλώματος επεξεργασίας ανταμοιβών. Ο γνωσιακός μηχανισμός εκτίμησης των απολαβών που συνδέεται με το κύκλωμα ανταμοιβών υπολογίζει την αξία μιας ατομικής απολαβής προσμετρώντας και τις απολαβές των λοιπών συμμετεχόντων σε μια αλληλεπίδραση, καθώς και τις προθέσεις δίκαιης συμπεριφοράς που αποκαλύπτουν οι επιλογές τους. Η παραβίαση συνεργατικών προτύπων φαίνεται ότι μειώνει την αντιλαμβανόμενη αξία μιας απολαβής, προκαλώντας αρνητικές αποκρίσεις του κυκλώματος ανταμοιβής. Τα αρνητικά αυτά συναισθήματα πυροδοτούν μια επιθυμία για τιμωρία του παραβάτη, η επιβολή της οποίας συνδέεται με θετικές αποκρίσεις του κυκλώματος ανταμοιβής. Ο μηχανισμός που συνδέει αυτά τα αρνητικά συναισθήματα θυμού με την πυροδότηση της επιθυμίας για τιμωρία και την προσδοκία θετικών συναισθημάτων από την επιβολή της φαίνεται ότι είναι εγγενής και συνδέεται με την αντίληψη της δικαιοσύνης στις ανθρώπινες αλληλεπιδράσεις. Οι κοινωνικές προτιμήσεις των ατόμων φαίνεται λοιπόν να αποτελούν ένα πολύ ισχυρό κίνητρο για τη συμπεριφορά τους και το γεγονός ότι υπάρχουν νευρωνικοί

μηχανισμοί που κρύβονται πίσω από την εκδήλωσή τους δείχνει ότι οι εξελικτικές τους ρίζες είναι πολύ βαθιές.

2.4 Εξελικτικά μοντέλα

Η αλτρουιστική τιμωρία φαίνεται να αποτελεί μια συμπεριφορά που βασίζεται σε μια παρορμητική αρνητική αντίδραση των ανθρώπων στην αδικία και τις παραβιάσεις των κοινωνικών κανόνων. Αυτά τα αισθήματα έντονης δυσαρέσκειας συνοδεύονται από μια ενεργητική διάθεση και τάση για επιβολή κυρώσεων στον παραβάτη. Η επιβολή αυτών των ποινών φαίνεται να προκαλεί στους τιμωρούς θετικά συναισθήματα και ευχαρίστηση, καθώς αισθάνονται ότι η δικαιοσύνη επανορθώνεται. Το αποτέλεσμα της επιβολής και της απειλής των ποινών είναι η μεγαλύτερη συμμόρφωση των ατόμων με τους συνεργατικούς κανόνες. Έτσι λοιπόν, στο εγγύς επίπεδο, το επίπεδο όπου λειτουργούν οι γνωσιακές διαδικασίες και εκδηλώνονται τα αποτελέσματά τους, εμφανίζεται ένα σύνολο κοινωνικών προτιμήσεων και κινήτρων, τα οποία διαφέρουν από τα εγωιστικά κίνητρα της κλασικής οικονομικής θεωρίας. Ο εγγενής χαρακτήρας των μηχανισμών που γεννούν αυτά τα κίνητρα μας ωθεί να αναζητήσουμε τις εξελικτικές τους ρίζες και τον τρόπο με τον οποίο μπόρεσαν να προσφέρουν κάποιο εξελικτικό πλεονέκτημα στους φορείς τους, προκειμένου να ευνοηθούν από τη φυσική επιλογή και να περάσουν στις επόμενες γενιές.

2.4.1 Πολιτισμική-γονιδιακή συνεξέλιξη

Ένα θεωρητικό μοντέλο που προσφέρει μια καλή εξήγηση για το πώς εξελικτικά ευνοήθηκε η εδραίωση της αλτρουιστικής τιμωρίας είναι αυτό της πολιτισμικής-γονιδιακής συνεξέλιξης (*gene-culture coevolution*) (Bowles et al., 2003; Boyd & Richerson, 2005) που συνδέεται σε μεγάλο βαθμό με το θεωρητικό μοντέλο της πολιτισμικής ομαδικής επιλογής (*cultural group selection*) (Gintis, 2000; Richerson et al., 2003). Η θεωρία αυτή πρεσβεύει ότι η ανάδυση ικανοτήτων πολιτισμικής μάθησης στη διάρκεια της ανθρώπινης εξέλιξης δημιούργησε διαδικασίες ομαδικής επιλογής που άλλαξαν τα επιλεκτικά περιβάλλοντα στα οποία αναπτύσσονται και προσαρμόζονται τα γονίδια. Για παράδειγμα, η πρακτική του μαγειρέματος του κρέατος μπορούμε να υποθέσουμε ότι διαδόθηκε μέσω της μιμητικής μάθησης σε αρχέγονους ανθρώπινους πληθυσμούς. Σε ένα περιβάλλον ‘μαγειρεμένου κρέατος’ η φυσική

επιλογή μπορεί να ευνόησε γονίδια που μείωναν το μήκος των ενεργοβόρων πεπτικών μας σωλήνων και άλλαζαν τους χημικούς μηχανισμού της πέψης μας. Μια τέτοια μείωση του πεπτικού ιστού μπορεί να απελευθέρωσε ενέργεια που συνέβαλλε στην αύξηση του εγκεφαλικού ιστού και των γνωσιακών ικανοτήτων. Κατά αυτό τον τρόπο, η ανθρώπινη βιολογία προσαρμόστηκε σε μια πολιτισμικά μεταδιδόμενη συμπεριφορά. Αυτή η αλληλεπίδραση ονομάζεται πολιτισμική-γονιδιακή συνεξέλιξη. Βάση αυτής της θεωρίας αποτελεί η υπόθεση ότι οι άνθρωποι διαθέτουν μια έμφυτη ικανότητα να μιμούνται και να εσωτερικοποιούν κανόνες κοινωνικής συμπεριφοράς.

Όταν μιλάμε για εσωτερικοποίηση κοινωνικών κανόνων (*internalization of norms*) εννοούμε την τήρηση του προτύπου συμπεριφοράς από το άτομο λόγω της ύπαρξης κάποιων εσωτερικών ποινών (ντροπή, τύψεις, απώλεια αυτοεκτίμησης), εν αντιθέσει με την τήρησή του λόγω της επιβολής εξωτερικών ποινών (Gintis, 2003). Η εσωτερικοποίηση των κανόνων αποτελεί τη γνωσιακή διαδικασία που δημιουργεί τις κοινωνικές προτιμήσεις στα άτομα. Ο λόγος που η δυνατότητα για εσωτερικοποίηση των κανόνων έχει εξελικτική αξία είναι ότι τα άτομα που έχουν αυτήν την τάση μπορούν να προσαρμοστούν ταχύτερα και αποτελεσματικότερα σε ένα ιδιαίτερα πολύπλοκο περιβάλλον. Οι αρχέγονες συνθήκες και η γρήγορη πολιτισμική εξέλιξη των ανθρώπων δημιουργούσαν ένα πολύ απαιτητικό εξελικτικό περιβάλλον για το ανθρώπινο είδος, γεμάτο με πολλές απειλές και προκλήσεις. Η προσαρμογή σε αυτό μπορούσε να γίνει πολύ πιο αποτελεσματικά αν κανείς ακολουθούσε πρότυπα συμπεριφοράς τα οποία έδειχναν επιτυχημένα και ακολουθούνταν ήδη από άλλα άτομα. Τα άτομα που διέθεταν ένα γνωσιακό μηχανισμό που τα έκανε να έχουν την τάση αυτή της μίμησης επιτυχημένων ή καθιερωμένων συμπεριφορών είχαν συνεπώς εξελικτικό πλεονέκτημα σε σχέση με τους υπόλοιπους. Η τάση για εσωτερικοποίηση κοινωνικών κανόνων επέτρεπε την πολιτισμική προσαρμογή, προσφέροντας έτσι τη δυνατότητα για ταχεία αύξηση της ατομικής *αρμοστικότητας (fitness)*, ενώ μια καθαρά γενετική προσαρμοστική διαδικασία θα διαρκούσε κατά τάξεις μεγέθους περισσότερο.

Η εσωτερικοποίηση κανόνων γενικά πραγματοποιείται μέσω της κοινωνικοποίησης (*socialisation*) και της κοινωνικής μάθησης (*social learning*) από τους γονείς (κάθετη μετάδοση-*vertical transmission*) ή από άλλους ανθρώπους (οριζόντια μετάδοση-*horizontal transmission*). Ένα μεγάλο κομμάτι της ανθρώπινης συμπεριφοράς οφείλεται στην κοινωνική μάθηση κι αυτό αποτελεί μια ειδοποιό διαφορά από τα υπόλοιπα πρωτεύοντα. (Henrich & Boyd, 1998; Takahasi,

1999; Harris, 1998). Ωστόσο, οι άνθρωποι δεν αντιγράφουν απλά τους γονείς ή τους συνανθρώπους τους τυχαία, αλλά φαίνεται ότι βάζουν σε εφαρμογή γνωσιακούς μηχανισμούς όπως η μίμηση αυτού που κάνει η πλειοψηφία, γνωστή ως μετάδοση βάσει συμμόρφωσης στα καθιερωμένα (*conformist based transmission*) και η μίμηση των επιτυχημένων ατόμων, γνωστή ως μετάδοση βάσει απολαβών ή κύρους (*payoff or prestige based transmission*). Αυτές οι εξειδικευμένες ευρετικές μέθοδοι (*heuristics*) δίνουν στα άτομα την ευκαιρία να παρακάμψουν τον κόπο και το χρόνο που απαιτεί η ατομική μάθηση και ο πειραματισμός και να κάνουν άλματα σε πιο προσαρμοστικές συμπεριφορές.

Από τη στιγμή που η εσωτερικοποίηση κοινωνικών κανόνων στους ανθρώπους ευνοείται εξελικτικά, η αλτρουιστική συμπεριφορά μπορεί να να εξηγηθεί ως ένα χαρακτηριστικό που ευνοήθηκε σαν παρελκόμενο αυτής της τάσης των ανθρώπων (Simon, 1990). Παρόλο δηλαδή που η αλτρουιστική συμπεριφορά έχει μεγαλύτερο υλικό κόστος απ' ότι όφελος για το άτομο, και άρα μειώνει την αρμοστικότητα του σε βιολογικούς όρους, η ακολούθηση των προτύπων αλτρουιστικής συμπεριφοράς από τους ανθρώπους μπορεί να ειδωθεί ως 'παρενέργεια' της γενικής τους τάσης να εσωτερικοποιούν κοινωνικούς κανόνες. Οι περισσότεροι κοινωνικοί κανόνες αυξάνουν την ατομική αρμοστικότητα. Η ακολούθηση κάποιων κανόνων που μειώνουν την αρμοστικότητα είναι λοιπόν πιθανή λόγω του γεγονότος αυτού, καθώς η τάση των ατόμων να εσωτερικοποιούν κανόνες δε μπορεί να πραγματοποιήσει το διαχωρισμό μεταξύ τους.

Έτσι, σε κάποια στενά χρονικά και γεωγραφικά εξελικτικά πλαίσια, μπορούμε να υποθέσουμε ότι ένας μειωτικός της αρμοστικότητας κανόνας είναι πιθανόν να αρχίσει να ακολουθείται από κάποια άτομα εντός μιας ομάδας. Για τον έλεγχο των υποθέσεων της εξελικτικής ανθρωπολογίας συχνά χρησιμοποιούνται μοντέλα προσομοίωσης των εξελικτικών διαδικασιών. Στα μοντέλα αυτά σε ένα υπολογιστή προσομοιώνονται αλληλεπιδράσεις μεταξύ ατόμων που φέρουν κάποια συγκεκριμένη συμπεριφορά. Η προσομοίωση τρέχει για πολλές εξελικτικές περιόδους και με μαθηματικές μεθόδους υπολογίζεται η σύσταση του πληθυσμού από φορείς των συμπεριφορών σε κάθε νέα περίοδο, προσμετρώντας την αρμοστικότητα που προσδίδει η καθεμιά από αυτές. Στη λήξη της προσομοίωσης βγαίνουν συμπεράσματα για τις υποθέσεις του μοντέλου με βάση την τελική σύσταση του πληθυσμού και την πιθανή επικράτηση των φορέων κάποιων συμπεριφοράς.

Τέτοια εξελικτικά μοντέλα έχουν δείξει ότι αν υπάρχει η δυνατότητα επιβολής τιμωρίας, οποιοσδήποτε κοινωνικός κανόνας μπορεί να σταθεροποιηθεί στα πλαίσια μιας ομάδας, ακόμα

και αν μειώνει το ατομικό όφελος και την αρμοστικότητα των μελών (Boyd & Richerson, 1992). Ο βασικός γνωσιακός μηχανισμός της εσωτερικοποίησης κανόνων στον οποίο μπορεί να βασιστεί το φαινόμενο αυτό είναι αυτός της μετάδοσης βάσει συμμόρφωσης στα καθιερωμένα, κάτι που έχει υλοποιηθεί σε μοντέλο εξελικτικής ανθρωπολογίας όπου τα άτομα έχουν τη δυνατότητα να τιμωρήσουν τους παραβάτες αλλά να τιμωρήσουν και αυτούς που απλά συνεργάζονται αλλά δεν τιμωρούν τους παραβάτες (τιμωρία δεύτερης τάξης, *second order punishment*).(Henrich & Boyd, 2001).

Οι γενικές αρχές του μοντέλου έχουν ως εξής: Λόγω της τάσης των ανθρώπων να μιμούνται και να μαθαίνουν τις συμπεριφορές που είναι πιο συχνές σε μια ομάδα, η επιβολή αλτρουιστικών τιμωριών από κάποια μέλη της ομάδας όταν παραβιάζονται συνεργατικά πρότυπα μπορεί να γίνει μια σταθερή συμπεριφορά για όλη την ομάδα. Όπως έχουμε δει και προηγουμένως από εμπειρικά πειραματικά δεδομένα, η ύπαρξη μιας κρίσιμης μάζας έστω και λίγων αλτρουιστικών τιμωρών μπορεί να προκαλέσει την αποθάρρυνση των λαθρεπιβατών και να τους παρασύρει όλους σε συνεργατικές επιλογές. Έτσι, αν σε μια ομάδα υπάρχουν ορισμένα άτομα που τιμωρούν τους λαθρεπιβάτες, τότε η μη συνεργατική συμπεριφορά δε συμφέρει, ειδικά αν οι ποινές που επιβάλλονται είναι σχετικά μεγάλες. Η συνεργατική συμπεριφορά λοιπόν γίνεται αντικείμενο μίμησης γιατί έχει μεγαλύτερες απολαβές και τα άτομα μιμούνται τις επιτυχημένες συμπεριφορές.

Σε αυτή την περίπτωση ο απολαβές των ατόμων που τιμωρούν (*punishers*) είναι σε μικρό βαθμό χαμηλότερες από αυτές των ατόμων που απλά συνεργάζονται χωρίς όμως να τιμωρούν τους παραβάτες (*contributors*), γιατί δεν υπάρχουν πολλοί λαθρεπιβάτες και άρα το κόστος τιμωρίας τους δεν είναι πολύ μεγάλο. Τα απλά συνεργατικά άτομα θεωρούνται λαθρεπιβάτες δεύτερης τάξης (*second order free-riders*), καθώς αποφεύγουν το κόστος επιβολής της τιμωρίας. Οι τιμωροί μπορούν να επιβάλουν ποινές και σε αυτά τα άτομα, κάτι που ονομάζεται τιμωρία δεύτερης τάξης (*second order punishment*). Οι τάξεις αυτές μπορούν να συνεχιστούν για πολλά επίπεδα και σε κάθε επίπεδο οι τιμωροί επιβάλλουν ποινές τόσο στους λαθρεπιβάτες όσο και στους απλά συνεργατικούς αυτής της τάξης.

Σε κάθε τάξη υπάρχει μια μικρή διαφορά των απολαβών μεταξύ των τιμωρών και των απλά συνεργατικών ατόμων, καθώς τα τελευταία αποφεύγουν τα κόστη επιβολής ποινής στους λαθρεπιβάτες. Όσο ανεβαίνουμε σε ανώτερες τάξεις τιμωρίας αυτή η διαφορά μεταξύ των απολαβών των τιμωρών και των απλά συνεργατικών ατόμων μειώνεται, μέχρι που σε σε κάποια

τάξη φτάνει να τείνει στο μηδέν, επειδή οι περιπτώσεις των παραβάσεων γίνονται όλο και λιγότερες. Όταν η διαφορά μεταξύ των απολαβών γίνει αρκετά μικρή και ουσιαστικά ασήμαντη, η τάση των ατόμων να μιμούνται τις πιο συχνές συμπεριφορές (μετάδοση βάσει συμμόρφωσης με τα καθιερωμένα) μπορεί να υπερισχύσει της τάσης τους να μιμούνται τις συμπεριφορές με τις μεγαλύτερες απολαβές (μετάδοση βάσει απολαβών) και έτσι μπορεί να ξεκινήσει η μίμηση του μειωτικού της αρμοστικότητας κανόνα.

Με τον τρόπο αυτό η αλτρουιστική τιμωρία, παρόλο που μειώνει την αρμοστικότητα σε υλικούς όρους, μπορεί να εδραιωθεί σε κάποια τάξη τιμωρίας. Άπαξ και η αλτρουιστική τιμωρία εδραιωθεί σε κάποια τάξη, εδραιώνεται στη συνέχεια και στην αμέσως επόμενη, αφού το να μην είσαι τιμωρός στην κατώτερη τάξη θα επιφέρει την τιμωρία των τιμωρών της ανώτερης τάξης και επομένως δεν είναι συμφέρον. Έτσι η τιμωρητική συμπεριφορά εδραιώνεται σε όλες τις τάξεις, μέχρι την πρώτη όπου τιμωρούνται τα άτομα που δε συνεργάζονται (λαθρεπιβάτες πρώτης τάξης). Με τον τρόπο αυτό, η συμπεριφορά της αλτρουιστικής τιμωρίας μέσα σε μια μεγάλη ομάδα μπορεί να σταθεροποιηθεί λόγω της τάσης των ατόμων να μιμούνται συχνές συμπεριφορές.

2.4.2 Πολιτισμική ομαδική επιλογή

Έτσι λοιπόν, παρόλο που η αλτρουιστική τιμωρία μειώνει την ατομική αρμοστικότητα, μπορεί να καθιερωθεί σα συμπεριφορά στα πλαίσια μιας ομάδας. Για να μπορεί όμως ένας κανόνας που μειώνει την ατομική αρμοστικότητα να επιμείνει διαχρονικά στη διαδικασία της εξέλιξης, θα πρέπει να υπάρχει κάποια αντίρροπη δύναμη που να αντισταθμίζει τις αρνητικές εξελικτικές πιέσεις της φυσικής επιλογής εναντίον των ατόμων που τον ακολουθούν εντός μιας ομάδας (*within group selection*). Υπάρχουν ενδείξεις ότι η δύναμη αυτή μπορεί να είναι η πολιτισμική ομαδική επιλογή (*cultural group selection*) (Gintis, 2000; Henrich & Boyd, 2001; Boyd et al., 2003).

Σύμφωνα με τη θεωρία αυτή, μια συμπεριφορά που σχετίζεται με πολιτισμικά χαρακτηριστικά μπορεί να επιμείνει σε μια ομάδα και να μεταδοθεί και σε άλλες, εφόσον αυξάνει την ομαδική αρμοστικότητα, ακόμα και αν μειώνει την ατομική αρμοστικότητα των μελών. Με αυτόν τον τρόπο η φυσική επιλογή μεταξύ των ομάδων (*between group selection*) ευνοεί τα άτομα με γνωρίσματα που στηρίζουν την εκδήλωση της συμπεριφοράς, αντισταθμίζοντας την αρνητική πίεση εναντίον τους λόγω της φυσικής επιλογής εντός της

ομάδας. Η πολιτισμική ομαδική επιλογή βασίζεται στις θεωρίες της ομαδικής επιλογής σε πολλαπλά επίπεδα (*multilevel group selection*) και την επεκτείνει για χαρακτηριστικά που μεταδίδονται πολιτισμικά (Sober & Wilson, 1998).

Η αλτρουιστική τιμωρία σταθεροποιεί και ισχυροποιεί τη συνεργασία εντός μιας ομάδας, προσφέροντας έτσι στην ομάδα μεγαλύτερη αρμοστικότητα και εξελικτικό πλεονέκτημα. Στο αρχέγονο εξελικτικό περιβάλλον αυτό ήταν πολύ σημαντικό, καθώς οι καταστάσεις που έπρεπε να αντιμετωπίσουν τα άτομα ήταν εξαιρετικά αβέβαιες. Οι έντονες φυσικές απειλές, οι λιμοί, οι φυσικές καταστροφές, ο πόλεμος μεταξύ φυλών ήταν πολύ συχνά περιστατικά που καθιστούσαν το μέλλον πολύ ασταθές. Η συνεργασία μεταξύ των ατόμων ήταν κρίσιμη για την επιβίωση και η ισχύς της αλτρουιστικής τιμωρίας ως σταθεροποιητικού παράγοντα της απαραίτητη, εφόσον η αβεβαιότητα δε θα επέτρεπε τη μακροημέρευση της αν βασιζόταν σε οποιαδήποτε άλλη μορφή αμοιβαιότητας. Έτσι λοιπόν οι ομάδες όπου υπήρχαν πολλοί αλτρουιστικοί τιμωροί είχαν και πιο ισχυρή συνεργασία, αποκτώντας εξελικτικό πλεονέκτημα έναντι των άλλων. Το πλεονέκτημα αυτό υπερίσχυε της αρνητικής εξελικτικής πίεσης που υφίσταντο οι τιμωροί εντός της ομάδας εξαιτίας της μείωσης του ατομικού τους οφέλους από το κόστος επιβολής των τιμωριών. Το γεγονός επίσης ότι η αλτρουιστική τιμωρία σαν συμπεριφορά βασίζεται στην εσωτερικοποίηση κανόνων και μεταδίδεται μεταξύ των ατόμων μέσω κοινωνικής μάθησης, της επέτρεπε να αντισταθεί σε εξωτερικές εισβολές από λαθρεπιβάτες και να επιβιώσει σε μια ομάδα.

Το τελικό αποτέλεσμα είναι ότι σε έναν πληθυσμό υπάρχουν τόσο άτομα που συμπεριφέρονται περισσότερο ως εγωιστικοί μεγιστοποιητές, όσο και άτομα που χαρακτηρίζονται κυρίως από συμπεριφορές αμοιβαιότητας. Η κάθε κατηγορία δέχεται ευνοϊκές εξελικτικές πιέσεις από αντίθετες κατευθύνσεις: οι εγωιστές ευνοούνται από τις εξελικτικές πιέσεις εντός της ομάδας, ενώ οι αλτρουιστές από τις εξελικτικές πιέσεις μεταξύ των ομάδων. Η ισορροπία ανάμεσα σε αυτές τις δύο κατηγορίες είναι διαρκώς ρευστή και καθορίζει τελικά και το βαθμό συνεργασίας μέσα σε μια ομάδα, καθώς όσο η κάθε κατηγορία αυξάνει σε αριθμό μπορεί να συμπαρασύρει προς το μέρος της την άλλη κατηγορία. Εξελικτικά μοντέλα που προσομοιώνουν τις παραπάνω υποθέσεις προσφέρουν ενδείξεις ότι η αλτρουιστική τιμωρία μπορεί να έχει εξελικτική σταθερότητα και να λειτουργήσει ως ενισχυτικός και σταθεροποιητικός παράγοντας για τη συνεργασία (Bowles & Gintis 2004; Boyd et al., 2003, Fowler, 2005; Sigmund et al., 2001).

Κεφάλαιο 3: Αλτρουιστική τιμωρία: εξωτερική εγκυρότητα

Τα ευρήματα των μελετών που περιγράφηκαν παραπάνω συνθέτουν ένα ισχυρό σώμα ενδείξεων για την ύπαρξη της τάσης για αλτρουιστική τιμωρία στους ανθρώπους, καθώς και για το θετικό ρόλο που μπορεί να παίξει για την ενίσχυση της συνεργασίας και των κοινωνικών κανόνων. Τα ευρήματα αυτά προέρχονται από συμπεριφορικά πειράματα της θεωρίας παιγνίων, από νευροαπεικονιστικές μελέτες και από εξελικτικά μοντέλα.

Η κατεύθυνση στην οποία συγκλίνουν τα ευρήματα είναι η επιβεβαίωση της παρουσίας κοινωνικών προτιμήσεων στη γκάμα των ανθρώπινων κινήτρων, οι οποίες χτίζονται στη βάση μιας γενετικής προδιάθεσης και ικανότητας των ανθρώπων για εσωτερικοποίηση κανόνων. Οι άνθρωποι μαθαίνουν να ακολουθούν αυτούς τους κανόνες και έχουν έντονες συναισθηματικές αντιδράσεις όταν αυτοί παραβιάζονται. Η συναισθηματική αυτή εμπλοκή επηρεάζει τη διαδικασία λήψης αποφάσεων και μπορεί να οδηγήσει σε πρότυπα συμπεριφοράς που αποκλίνουν από το κλασικό εγωιστικό μοντέλο της θεωρίας της χρησιμότητας όπου όλες οι επιλογές έχουν σα στόχο τη μεγιστοποίηση του υλικού οφέλους του ατόμου.

Η αλτρουιστική τιμωρία αποτελεί μια τέτοια απόκλιση από το εγωιστικό μοντέλο και μπορεί να ειδωθεί σα μια εκδήλωση των επιπτώσεων των αρνητικών συναισθημάτων που προκαλεί στα άτομα η παραβίαση των κοινωνικών κανόνων από άλλα μέλη της ομάδας. Η απειλή της εφαρμογής της αποθαρρύνει τους παραβάτες και οδηγεί σε ενίσχυση της συνεργασίας και σε εδραίωσή της. Το εξελικτικό μειονέκτημα που μπορεί να επιφέρει αυτού του είδους η συμπεριφορά σε ατομικό επίπεδο μπορεί να αντισταθμιστεί από το πλεονέκτημα που προσφέρει σε επίπεδο ομάδας, κάτι που την καθιστά εξελικτικά σταθερή. Έτσι λοιπόν μπορούμε να έχουμε μια βάση για την ανάπτυξη και τη διατήρηση συνεργατικών συμπεριφορών, η οποία μπορεί να εξηγήσει τη συνεργασία ακόμα και σε συνθήκες όπου οι αλληλεπιδράσεις μεταξύ των ατόμων δεν είναι επαναλαμβανόμενες, οι ομάδες είναι μεγάλες και δεν υπάρχουν δυνατότητες δημιουργίας φήμης. Η ισχυρή αμοιβαιότητα και οι σχετικοί με αυτή υποβόσκοντες γνωσιακοί μηχανισμοί λοιπόν μπορούν να δώσουν εξηγήσεις για φαινόμενα τα οποία δε μπορούν να εξηγηθούν επαρκώς μέσω της θεωρίας της άμεσης ή έμμεσης αμοιβαιότητας.

Όπως έχει ήδη αναφερθεί, σκοπός των πειραμάτων γύρω από την αλτρουιστική τιμωρία είναι να απομονωθούν σε περιβάλλον εργαστηρίου τα κίνητρα που βρίσκονται πίσω από τη συνεργατική συμπεριφορά. Η μέθοδος αυτή είναι απαραίτητη αν θέλουμε να διαχωρίσουμε τα

όποια συναισθηματικά κίνητρα συνεργασίας έχουν τα άτομα λόγω των κοινωνικών τους προτιμήσεων από τα εγωιστικά τους κίνητρα που αφορούν την αύξηση του ατομικού τους οφέλους μέσω της συνεργασίας. Με τον έλεγχο των παραμέτρων (επαναλήψεις αλληλεπιδράσεων, ανωνυμία) στο εργαστήριο έχουμε τη δυνατότητα να αποκαλύψουμε μια κατηγορία κινήτρων που διαφορετικά θα παρέμενε συγκαλυμμένη από τις εγωιστικά κίνητρα αμοιβαιότητας. Το πλεονέκτημα όμως αυτό μετατρέπεται σε μειονέκτημα για την εξαγωγή συμπερασμάτων για τον πραγματικό κόσμο. Στην πραγματική ζωή οι συνθήκες αλληλεπίδρασης μεταξύ των ατόμων διαφέρουν από αυτές του εργαστηρίου και το όποιο συμπέρασμα πρέπει να ελεγχθεί με στοιχεία συγκεντρωμένα από εμπειρική παρατήρηση πραγματικών καταστάσεων. Τα στοιχεία αυτά θα πρέπει να συγκλίνουν με τα αποτελέσματα των πειραμάτων προκειμένου αυτά να έχουν τη λεγόμενη εξωτερική εγκυρότητα (*external validity*), να επιβεβαιώνονται δηλαδή στον πραγματικό κόσμο (Starmer, 1999; Guala, 2005; Bardsley et al., 2009).

Έτσι, παρά το γεγονός ότι τα αποτελέσματα έχουν αναπαραχθεί πολλές φορές σε διάφορες πειραματικές παραλλαγές, έχει διατυπωθεί η αμφισβήτηση ότι η συμπεριφορά που διαπιστώνεται στα πειράματα δεν αποτελεί μια αυθεντική αντίδραση των ατόμων στις δεδομένες συνθήκες, αλλά μεταφορά μαθημένων στρατηγικών που χρησιμοποιούν συχνά στον πραγματικό κόσμο. Σε συνθήκες πραγματικής ζωής, όπου συνήθως οι αλληλεπιδράσεις είναι επαναλαμβανόμενες και υπάρχει πληροφορία για την ταυτότητα και το ιστορικό των αλληλεπιδρώντων, οι στρατηγικές αυτές είναι επιτυχημένες και για αυτό μετατρέπονται συχνά σε ευρετικές μεθόδους (Binmore, 1999; Binmore, 2006; Trivers, 2004; Burnham & Johnson, 2005). Όπως έχει ήδη αναφερθεί ο ισχυρισμός αυτός δεν ευσταθεί, εφόσον όπως ήδη έχει δειχθεί στα πειράματα που παρουσιάστηκαν τα άτομα είναι σε θέση να διαχωρίσουν τις διαφορετικές συνθήκες και τις επιπλοκές τους και να προσαρμόσουν ανάλογα τη συμπεριφορά τους (Fehr & Fischbacher, 2003; Fehr & Gächter, 2000).

3.1 Απόκλιση πειραματικών συνθηκών από τις πραγματικές

Το βασικό λοιπόν πρόβλημα δεν είναι ότι τα άτομα δεν κατανοούν τις πειραματικές συνθήκες και ακολουθούν ευρετικές μεταφερμένες από τον πραγματικό κόσμο, αλλά αντιθέτως ότι οι συνθήκες αυτές είναι δυνατόν να ευνοούν την εμφάνιση μιας συμπεριφοράς που στα πλαίσια της πραγματικής ζωής δε θα εκδηλωνόταν. Η εμφάνιση της συμπεριφοράς μπορεί επομένως να

αποτελεί ένδειξη μιας τάσης, ενός κινήτρου ή μιας επιθυμίας των ατόμων που βρίσκεται πίσω από αυτή, ωστόσο η ίδια η συμπεριφορά δε μπορούμε να θεωρήσουμε ότι έχει οικολογική εγκυρότητα. Για να το κάνουμε αυτό θα πρέπει να έχουμε πρόσθετες ενδείξεις από ανθρωπολογικές παρατηρήσεις κοινωνιών όπου παρόμοια συμπεριφορά εκδηλώνεται σε πραγματικές συνθήκες παρόμοιες με αυτές του πειράματος.

Στην περίπτωση της αλτρουιστικής τιμωρίας κάτι τέτοιο είναι πολύ δύσκολο. Στην πραγματική ζωή οι αλληλεπιδράσεις μεταξύ των ατόμων λαμβάνουν χώρα σε ένα πλαίσιο που έχει σημαντικές διαφορές από αυτό του εργαστηρίου. Τα άτομα σε πραγματικές συνθήκες έχουν κάποιες πολύ σημαντικές δυνατότητες που ο σχεδιασμός των πειραμάτων περιορίζει: τη δυνατότητα να επικοινωνήσουν και να ανταλλάξουν πληροφορίες μεταξύ τους, τη δυνατότητα να επιλέξουν εναλλακτικές μορφές τιμωρίας, τη δυνατότητα να αποσυρθούν τελείως από την ομάδα και τη δυνατότητα να ανταποδώσουν την τιμωρία που τους έχει επιβληθεί. Καθεμία από αυτές τις δυνατότητες μπορεί να αλλάξει τελείως τα χαρακτηριστικά και το σκηνικό της αλληλεπίδρασης και να επηρεάσει τις επιλογές για επιβολή τιμωρίας στα μη συνεργατικά μέλη.

3.1.1 Δυνατότητα επικοινωνίας

Η δυνατότητα επικοινωνίας μεταξύ των ατόμων έχει πολύ μεγάλη σημασία για την επιβολή τιμωρίας επειδή μπορεί να επιτρέψει την πολύ μεγάλη μείωση του κόστους της τιμωρίας. Τα άτομα μπορούν να συνεννοηθούν μεταξύ τους για το πως θα μπορούσαν να μοιράσουν τα κόστη τιμωρίας των παραβατών, δημιουργώντας έτσι συμμαχίες ή συνασπισμούς, επιφορτισμένους με τη δικαιοδοσία να επιβάλλουν τους κοινωνικούς κανόνες (Ostrom, 1990). Μέσα από τα πειράματα γίνεται φανερό ότι το κόστος της επιβολής της αλτρουιστικής τιμωρίας είναι κάτι που τα άτομα προσμετρούν και συνυπολογίζουν στην απόφασή τους για το ύψος που αυτή θα έχει (Ledyard, 1995; Isaac & Walker, 1988; Egas & Riedl, 2008). Το γεγονός αυτό φαίνεται χαρακτηριστικά και στη νευροαπεικονιστική μελέτη των de Quervain et al. (2004) που περιγράφηκε παραπάνω, όπου τα άτομα εξαντλούν τα περιθώρια της τιμωρίας όταν η επιβολή της είναι δωρεάν, αλλά μετριάζουν το ύψος της ποινής όταν πρέπει να επωμιστούν κόστος για να την επιβάλουν. Ακόμα και όταν η επιθυμία για επιβολή τιμωρίας είναι έντονη, τα άτομα τη μετριάζουν ή και την αποφεύγουν εντελώς όταν το κόστος της αυξάνει ή υπερβαίνει κάποια όρια. Τα άτομα θέλουν να επιβάλουν τιμωρία και να αποκαταστήσουν τα πρότυπα των κοινωνικών κανόνων, ωστόσο πάντοτε επιχειρούν να μειώσουν το κόστος της. Η δυνατότητα για

συνεννόηση προσφέρει ένα τρόπο να γίνει αυτό και μεταβάλλει το χαρακτήρα επιβολής της αλτρουιστικής τιμωρίας. Τα άτομα δε χρειάζεται να επιβάλουν αλτρουιστική τιμωρία το καθένα ξεχωριστά αλλά μπορούν να επιλέξουν να το κάνουν από κοινού, μέσω θεσμών κοινών πόρων (*common pool resources*), επιμερίζοντας και μειώνοντας έτσι το κόστος και κανονικοποιώντας και νομιμοποιώντας την τιμωρία (Ostrom et al., 1994).

3.1.2 Μορφή τιμωρίας

Η μορφή της τιμωρίας που μπορεί να επιβληθεί έχει επίσης άμεση σχέση με τη μείωση του κόστους και τη δυνατότητα για κυκλοφορία της πληροφορίας σχετικά με τα μέλη μιας ομάδας και τις πράξεις τους. Οι ανθρώπινες κοινωνίες βασίζονται σε δίκτυα μετάδοσης πληροφορίας γύρω από τα άτομα και το ποιόν τους και η πληροφορία αυτή μπορεί να καθορίσει πολλές πτυχές της ζωής ενός ατόμου. Η μετάδοση αρνητικών σχολίων με τη μορφή του κουτσομπολιού μπορεί να βλάψει πάρα πολύ ένα άτομο, αν και φέρει πολύ μικρό κόστος για το άτομο που αναλαμβάνει την πρωτοβουλία. Ανθρωπολογικές μελέτες (Henrich & Henrich, 2007, ch. 7) δείχνουν ότι τα άτομα συχνά επιλέγουν τέτοιες έμμεσες και σχεδόν χωρίς κόστος μεθόδους για την επιβολή της τιμωρίας των παραβατών. Στον πραγματικό κόσμο η κατά πρόσωπο αντιμετώπιση και τιμωρία δεν είναι τόσο συχνό φαινόμενο. Τα άτομα δηλαδή προτιμούν να χρησιμοποιήσουν τον κοινωνικό ιστό προκειμένου να επιβάλουν μια έμμεση, άυλη τιμωρία στον παραβάτη η οποία δεν έχει ουσιαστικό υλικό κόστος για τα ίδια αλλά δεν αφαιρεί πόρους ούτε από τον τιμωρούμενο. Παρόλα αυτά είναι εξαιρετικά αποτελεσματική και συμβάλλει σε σημαντικό βαθμό στην τήρηση των κανόνων εντός μιας ομάδας. Αυτή η μορφή τιμωρίας ξεφεύγει από τον ορισμό της αλτρουιστικής τιμωρίας, εφόσον το μηδενικό σχεδόν κόστος της την καθιστά περισσότερο συμβολική παρά υλική. Η ευρύτητα της χρήσης και της σημασίας της συμβολικής τιμωρίας έχει φανεί και σε πειράματα, όπου παρατηρείται ότι τα άτομα είναι διατεθειμένα να επωμιστούν ακόμα και ένα μικρό χρηματικό κόστος προκειμένου να την επιβάλουν (Carpenter et al., 2004). Είναι λοιπόν εμφανές ότι ο περιορισμός των μορφών της τιμωρίας που μπορεί να επιβάλουν τα άτομα μέσω του πειραματικού σχεδιασμού μπορεί να επηρεάσει τα ευρήματα και να αυξήσει τα επίπεδα της αλτρουιστικής τιμωρίας, παρόλο που στον πραγματικό κόσμο τα άτομα θα επέλεγαν κάποια άλλη μορφή ποινής.

3.1.3 Δυνατότητα αποχώρησης από το παιχνίδι

Πέρα από τις διαφορετικές μορφές τιμωρίας που μπορούν να επιλέξουν τα άτομα, σημαντική επίδραση στα πειραματικά αποτελέσματα μπορεί να έχει και το γεγονός ότι στα άτομα δε δίνεται η δυνατότητα να αποχωρίσουν από το παιχνίδι. Στην πραγματική ζωή τα άτομα έχουν το περιθώριο να επιλέξουν με ποιον θα αλληλεπιδράσουν και δεν είναι υποχρεωμένα να παραμείνουν στα πλαίσια μιας συγκεκριμένης ομάδας. Οι επιλογές τους δηλαδή δεν περιορίζονται στο να συνεργαστούν ή να μη συνεργαστούν αλλά μπορούν να επιλέξουν επίσης να μη συμμετέχουν καθόλου στην αλληλεπίδραση. Γίνεται λοιπόν αντιληπτό ότι η δυσαρέσκεια των ατόμων όταν συμμετέχουν σε μια ομάδα όπου νιώθουν ότι οι κοινωνικοί κανόνες παραβιάζονται και υπάρχει αδικία μπορεί να εκδηλωθεί με διάφορους τρόπους στην πραγματική ζωή. Τα συμπεριφορικά πειράματα μέσω του σχεδιασμού τους περιορίζουν τα κανάλια μέσα από τα οποία μπορεί να εκδηλωθεί αυτή η δυσαρέσκεια, προσφέροντας στα άτομα μόνο την επιλογή της τιμωρίας με υλικό κόστος. Με τον τρόπο αυτό είναι πιθανό να τονίζουν το φαινόμενο της αλτρουιστικής τιμωρίας και να του δίνουν διαστάσεις που δεν έχει στην πραγματική ζωή. Σε πειράματα μάλιστα που έχουν γίνει έχει βρεθεί ότι όταν στα άτομα δίνεται η δυνατότητα να επιλέξουν, τότε διαλέγουν να στηρίξουν τη συνεργασία με ένα μίγμα συμβολικής τιμωρίας, άμεσης αμοιβαιότητας και, τέλος, τιμωρίας που η επιβολή της απαιτεί υλικό κόστος (Ostrom et al., 1992; Xiao & Houser, 2005; Ule et al., 2009).

3.1.4 Δυνατότητα αντι-τιμωρίας

Μια ακόμα πολύ σημαντική διαφορά του πραγματικού κόσμου από τα πειράματα είναι ότι ο σχεδιασμός των πειραμάτων δεν επιτρέπει στους τιμωρούμενους να αντιδράσουν στην τιμωρία που δέχονται. Δε μπορούν δηλαδή τα άτομα να ανταποδώσουν την τιμωρία που τους επιβάλλεται. Το σημείο αυτό είναι εξαιρετικά κρίσιμο. Αν οι τιμωροί γνωρίζουν ότι η τιμωρία ενέχει τον κίνδυνο αντιποίνων από τον τιμωρούμενο μπορεί να είναι πολύ λιγότερο διατεθειμένοι να προχωρήσουν στην επιβολή της. Σε περίπτωση επιβολής οι τιμωρούμενοι μπορεί να απαντήσουν με αντεκδίκηση και να δημιουργηθεί έτσι ένας φαύλος κύκλος τιμωρίας, ο οποίος φυσικά θα είναι καταστροφικός τόσο για τη συνεργασία όσο και για το συνολικό όφελος της ομάδας (Nikiforakis, 2008). Το αποτέλεσμα θα είναι η πλήρης κατασπατάληση και καταστροφή των πόρων, που θα θυσιαστούν στο βωμό μιας μορφής βεντέτας που θα επικρατήσει στην αλληλεπίδραση των ατόμων. Είναι κατανοητό λοιπόν ότι τα άτομα θα θέλουν να αποφύγουν κάτι τέτοιο και για το λόγο αυτό η επιβολή τιμωρίας μπορεί να περιοριστεί.

Πράγματι, σε πειράματα που περιλαμβάνουν τη δυνατότητα γα επιβολή αντι-τιμωρίας, βρέθηκε ότι οι συμμετέχοντες ήταν απρόθυμοι να επιβάλουν αλτρουιστική τιμωρία, δείχνοντας μια προτίμηση σε πιο ήπιες και 'φθηνές' στρατηγικές, όπως η απόσυρση της συνεργασίας (Nikiforakis & Engelmann, 2010; Dreber et al., 2008).

3.1.5 Κόστος τιμωρίας και συνολικές απολαβές

Τέλος, κάτι το οποίο συχνά παραβλέπεται στους πειραματικούς σχεδιασμούς και τις προσομοιώσεις είναι η γενική επίδραση που μπορεί να έχει το κόστος της αλτρουιστικής τιμωρίας στη συνεργασία μακροχρόνια. Η επιβολή αλτρουιστικής τιμωρίας, ακόμα και χωρίς τις αρνητικές επιπτώσεις της αντι-τιμωρίας, καταστρέφει πόρους χρήσιμους για την ομάδα εφόσον έχει κόστος τόσο για τον τιμωρό όσο και για τον τιμωρούμενο (Guala, 2012; Egas & Riedl, 1998). Αυτό φυσικά είναι εξαιρετικά επιβλαβές για μια ομάδα. Οι πόροι γενικά βρίσκονται σε στενότητα και το όφελος που προκύπτει από την αυξημένη συνεργασία που υποκινεί η τιμωρία μπορεί να μην είναι σε θέση να αντισταθμίσει τους πόρους που θυσιάζονται από την επιβολή της. Έτσι, ο ρόλος της αλτρουιστικής τιμωρίας για την υποστήριξη της συνεργασίας έχει σημαντικούς περιορισμούς στο απώτατο εξελικτικό επίπεδο.

3.2 Περιορισμοί της εφαρμογής και του ρόλου της αλτρουιστικής τιμωρίας

3.2.1 Περιορισμοί στο εγγύς και απώτατο επίπεδο

Είναι εμφανές ότι οι περιορισμοί που θέτουν οι πειραματικοί σχεδιασμοί κάνουν τις συνθήκες των πειραμάτων να αποκλίνουν σε σημαντικά σημεία από τις πραγματικές δημιουργώντας πρόβλημα για την εξωτερική εγκυρότητα των ευρημάτων και την προβολή των συμπερασμάτων στον πραγματικό κόσμο. Στην πραγματικές συνθήκες η αλτρουιστική τιμωρία και ο ρόλος της για τη συντήρηση της συνεργασίας δέχεται περιορισμούς τόσο στο εγγύς όσο και στο εξελικτικό επίπεδο. Ο βασικός της περιορισμός προκαλείται από τα υψηλά κόστη που επιφέρει. Στο εγγύς επίπεδο, τα άτομα υποκινούνται ταυτόχρονα από κίνητρα κοινωνικών προτιμήσεων και από εγωιστικά κίνητρα. Η αλτρουιστική τιμωρία υποκινείται από τις κοινωνικές προτιμήσεις και για να επιλέξουν να την επιβάλουν τα άτομα ζυγίζουν κόστη και οφέλη. Έτσι, μαζί με το συναισθηματικό όφελος που προκύπτει από την τιμωρία του παραβάτη συνυπολογίζεται και το

κόστος της επιβολής για τον τιμωρό, καθώς και η αναλογία του με τις επιπτώσεις στις απολαβές του τιμωρούμενου (Egas & Riedl, 1998). Αν η αναλογία δεν είναι ικανοποιητική ή το κόστος είναι πολύ μεγάλο, η αλτρουιστική τιμωρία αποφεύγεται. Σε ένα ενδιάμεσο και απώτατο επίπεδο επίσης, η επιβολή υψηλών ποινών μπορεί να οδηγήσει σε μεγάλη σπατάλη πόρων λόγω του κόστους για τιμωρό και τιμωρούμενο, μειώνοντας έτσι τις συνολικές απολαβές της ομάδας. Λόγω της ύπαρξης αντι-τιμωρίας το φαινόμενο της καταστροφής των πόρων εντείνεται και είναι πιθανό οι ομάδες να οδηγηθούν σε πλήρη κατάρρευση της συνεργασίας (Nikiforakis & Engelmann, 2010). Στον πραγματικό κόσμο επομένως οι άνθρωποι αναζητούν τρόπους να επιβάλλουν τους κοινωνικούς κανόνες και να 'διορθώσουν' τις παραβάσεις, μειώνοντας όσο μπορούν το κόστος.

3.2.2 Θεσμοί κοινών πόρων (common pool resources)

Όπως ήδη αναφέρθηκε ένας βασικός μηχανισμός μέσω του οποίου επιτυγχάνεται η μείωση του κόστους της τιμωρίας είναι η δημιουργία συμμαχιών όπου τα άτομα δρουν συντονισμένα και σε συνεννόηση προκειμένου να στοχεύσουν αποτελεσματικά την τιμωρία και να επιμερίσουν τα κόστη της (Casari, 2007; Ostrom, 1990; Ostrom, 2000). Η επικοινωνία μεταξύ των ατόμων παίζει εδώ βασικό ρόλο. Η ανάθεση της τιμωρίας σε μια συμμαχία ατόμων αποτελεί ένα πρώτο βήμα προς τη δημιουργία κεντρικών θεσμών για την επιβολή ποινών. Απομακρυνόμαστε δηλαδή από το πρότυπο της ατομικής, αυθαίρετης και μη συντονισμένης τιμωρίας που υλοποιείται στα πειράματα και περνάμε σε μια συντονισμένη μορφή της. Η συμμαχία και όχι το κάθε άτομο ξεχωριστά έχει πλέον το ρόλο και τη δικαιοδοσία του τιμωρού. Σε αυτήν παραχωρούν τα άτομα το δικαίωμα της επιβολής ποινών και την υποχρέωση για τον έλεγχο της εφαρμογής των κοινωνικών κανόνων. Η ροή των πληροφοριών για τις πράξεις και τη συμπεριφορά των ατόμων της ομάδας είναι πλέον το στοιχείο στο οποίο βασίζεται η συμμαχία για να επιτελέσει το έργο της.

Ένα πρόσφατο εξελικτικό μοντέλο ενσωμάτωσε αυτά τα νέα στοιχεία θέτοντας στις παραμέτρους τη δυνατότητα για επικοινωνία και συνασπισμό των ατόμων και το κόστος τιμωρίας αντιστρόφως ανάλογο του αριθμού των τιμωρών (Boyd et al., 2010). Το αποτέλεσμα της προσομοίωσης έδειξε ότι η συνεργασία μπορεί να σταθεροποιηθεί στη βάση της συντονισμένης τιμωρίας και μάλιστα τα συνολικά οφέλη για την ομάδα είναι αυξημένα σε σχέση με τα μοντέλα όπου η τιμωρία είναι ασυντόνιστη. Στην ίδια κατεύθυνση δείχνουν και τα

αποτελέσματα ενός πειράματος δημόσιου αγαθού, όπου μεταξύ των συμμετεχόντων παρουσιάστηκαν υψηλότερα επίπεδα συνεργασίας όταν η απόφαση για τιμωρία λαμβανόταν από ένα συνασπισμό σε σχέση με τη λήψη της απόφασης από μεμονωμένα άτομα (Casari & Luini, 2009).

Αυτή η παραχώρηση των δικαιωμάτων τιμωρίας στη συμμαχία των μελών της ομάδας είναι πιθανόν και η βάση για τη θεσμοποίηση και την συγκέντρωση των μηχανισμών τιμωρίας που παρατηρείται στις σύγχρονες κοινωνίες. Για σημαντικές παραβιάσεις των κοινωνικών κανόνων, η τιμωρία των οποίων θα απαιτούσε την ανάληψη ενός υψηλού ατομικού κόστους, υπεύθυνοι για την επιβολή ποινών είναι θεσμοί αστυνόμευσης και επιβολής δικαιοσύνης. Οι κοινωνίες έχουν παραχωρήσει σε θεσμούς κοινών πόρων (*common pool resources*) το ρόλο του τιμωρού, νομιμοποιώντας τους να επιβάλλουν ποινές στους παραβάτες των κοινωνικών κανόνων. Η συναίνεση αυτή καθιστά την τιμωρία αποδεκτή, κατανοητή και αναμενόμενη από το σύνολο των ατόμων, ακόμα και από τους τιμωρούμενους, απομακρύνοντας τον κίνδυνο της αντι-τιμωρίας. Έτσι η κοινωνία εξασφαλίζει μια σταθερή βάση για τη συνεργασία, μειώνοντας το κόστος της τιμωρίας και εμποδίζοντας τη διολίσθηση σε συνθήκες αντεκδίκησης που θα οδηγούσαν στην κατάρρευσή της. Πειράματα μάλιστα έχουν δείξει ότι τα άτομα προτιμούν πολλές φορές να μετακινηθούν σε κοινωνίες όπου υπάρχουν θεσμοί για την επιβολή συντονισμένης τιμωρίας. (Yamagishi, 1986, Gürerk et al., 2006).

3.2.3 Αλτρουιστική ανταμοιβή

Τέλος, είναι σημαντικό να αναφέρουμε ότι στην προσπάθεια για εξήγηση της συνεργασίας σπουδαίο ρόλο μπορεί να έχει παίξει και το θετικό στοιχείο της θεωρίας της ισχυρής αμοιβαιότητας, η αλτρουιστική ανταμοιβή. Πέρα από την προθυμία για ανάληψη κόστους για την τιμωρία παραβατών και λαθρεπιβατών, τα άτομα παρουσιάζουν επίσης προθυμία να αναλάβουν κόστος για να επιβραβεύσουν μια σύμφωνη με τους κανόνες συμπεριφορά. Ενδείξεις από μελέτες (Andreoni et al., 1998; Fong, 2001) και πειράματα (Berg et al., 1995; Fehr et al., 1993; Fischbacher et al., 2001) υποστηρίζουν την ύπαρξη αυτού του μηχανισμού κινήτρων Μάλιστα το γεγονός ότι τα άτομα έχουν διαφορετική υποκειμενική εκτίμηση ανάμεσα σε κέρδη και ζημίες (Kahneman et al., 1991) πιθανόν να αποτελεί ένδειξη ότι υπάρχει διαφοροποίηση των γνωσιακών μηχανισμών για την επεξεργασία των θετικών από τις αρνητικές ανταμοιβές. Η

ενίσχυση θετικών προτύπων μέσω του θετικού παραδείγματος και της επιβράβευσης σε παλαιότερες έρευνες έχει δείξει ότι μπορεί να στηρίξει τη μίμηση και ακολούθηση συνεργατικών προτύπων και αλτρουιστικών συμπεριφορών, όπως η βοήθεια ξένων σε μη επαναλαμβανόμενα πλαίσια (Bryan, 1971; Grusec, 1971). Η θετική ενίσχυση των πράξεων που είναι σύμφωνες με τους κοινωνικούς κανόνες μπορεί να αποτελεί ένα πολύ σημαντικό μηχανισμό για τη σταθεροποίηση της συνεργασίας, ειδικά αν αναλογιστεί κανείς πόσο ενισχυτική μπορεί να είναι η απλά λεκτική και συμβολική επιβράβευση, η οποία μπορεί να εφαρμοστεί με μηδενικό κόστος.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Το πρόβλημα της συνεργασίας αποτελεί ένα κεντρικό πρόβλημα των θεωριών για την ανάπτυξη της ανθρώπινης κοινωνικότητας που έχει απασχολήσει τους επιστήμονες για πολλά χρόνια. Οι θεωρίες που έχουν αναπτυχθεί περιλαμβάνουν την επιλογή βάσει γενετικής συγγένειας, την άμεση και την έμμεση αμοιβαιότητα και την ισχυρή αμοιβαιότητα. Ανάμεσα σε αυτές, η θεωρία της ισχυρής αμοιβαιότητας είναι η μόνη που είναι σε θέση να εξηγήσει τη συνεργασία σε μεγάλες ομάδες, όταν οι αλληλεπιδράσεις δεν είναι επαναλαμβανόμενες και όταν η πληροφορία για τα μέλη της ομάδας είναι φτωχή. Στοιχείο κλειδί για τη θεωρία της ισχυρής αμοιβαιότητας είναι η αλτρουιστική τιμωρία, η τάση δηλαδή που παρουσιάζουν οι άνθρωποι να τιμωρούν με κόστος τα άτομα που παραβαίνουν τους κανόνες κοινωνικής συμπεριφοράς, ακόμα και όταν αυτό δεν τους αποφέρει κάποιο άμεσο ή έμμεσο όφελος.

Η αλτρουιστική τιμωρία συνδέεται άμεσα με τις κοινωνικές προτιμήσεις των ατόμων. Οι προτιμήσεις αυτές ξεφεύγουν από τα στενά όρια των εγωιστικών προτιμήσεων για μεγιστοποίηση του ατομικού οφέλους και δείχνουν ότι τα άτομα ενδιαφέρονται για τις απολαβές των άλλων πέρα από τις δικές τους. Πέρα δηλαδή από το στενό τους οικονομικό όφελος σε υλικούς όρους, τα άτομα στην εσωτερική τους αντίληψη χρησιμότητας συνυπολογίζουν και την κατανομή των οφελών και στα άλλα μέλη μιας αλληλεπίδρασης, καθώς και τις προθέσεις που αυτά επιδεικνύουν. Είναι σαφής η προτίμηση των ατόμων για κατανομές απολαβών που βρίσκονται κοντά σε ένα πρότυπο δικαιοσύνης και ισότητας και η όποια απόκλιση από αυτό πρέπει να δικαιολογείται από τις επιμέρους συνθήκες για να είναι αποδεκτή.

Ένα ευρύ σώμα συμπεριφορικών πειραματικών δεδομένων αποκαλύπτει μια εγγενή επιθυμία και προδιάθεση των ατόμων να υπερασπιστούν τα κοινωνικά πρότυπα και να αντιδράσουν με επιβολή ποινών όταν αυτά καταστρατηγούνται. Ευρήματα από νευροαπεικονιστικές μελέτες προσφέρουν περαιτέρω στήριξη σε αυτά τα δεδομένα, δίνοντας ενδείξεις για την ύπαρξη ενός νευρωνικού μηχανισμού επεξεργασίας των απολαβών που συνδέεται με συναισθηματικές αντιδράσεις. Όταν τα άτομα αντιλαμβάνονται ότι αδικούνται ο μηχανισμός αυτός τα κάνει να αισθάνονται δυσαρέσκεια και οργή, η οποία μπορεί να εκδηλωθεί με τη μορφή της αλτρουιστικής τιμωρίας. Η επιβολή ποινής στους παραβάτες συνδέεται με την πρόκληση θετικής ανταμοιβής και αισθημάτων επανόρθωσης της δικαιοσύνης στους τιμωρούς.

Στο εγγύς επίπεδο λοιπόν, η αλτρουιστική τιμωρία μπορεί να ενταχθεί σε ένα σύνολο κινήτρων που γεννούν οι κοινωνικές προτιμήσεις των ατόμων. Η ύπαρξή της είναι αποτέλεσμα της επιθυμίας για τιμωρία της αδικίας, η οποία έχει γενετική βάση. Η γενετική προδιάθεση για ανάδυση συναισθημάτων θυμού και αντεκδίκησης όταν τα άτομα νιώθουν εξαπατημένα βασίζεται σε εγγενείς γνωσιακούς μηχανισμούς επεξεργασίας ανταμοιβών και συνδυάζεται με πρότυπα συμπεριφοράς και αντίδρασης στην αδικία που μεταδίδονται μέσω της κοινωνικής μάθησης και βασίζονται σε γνωσιακούς μηχανισμούς εσωτερικοποίησης κοινωνικών κανόνων. Πρόκειται για μια συναισθηματική πραγματικότητα την οποία βιώνει το σύνολο των ατόμων και ο τρόπος και η σφοδρότητα με την οποία μπορεί να εκδηλωθεί εξαρτάται επίσης από παράγοντες κοινωνικής μάθησης.

Παρόλα αυτά, η τάση για αλτρουιστική τιμωρία και η απειλή που αυτή επιφέρει παραμένει μία μόνο από τις συνιστώσες που κινούν τη συνεργατική συμπεριφορά. Οι πειραματικές μελέτες είναι απαραίτητες και χρήσιμες προκειμένου να τη διαχωρίσουμε και να τη φέρουμε στο φως, ωστόσο δεν είναι η μόνη ούτε απαραίτητα η βασικότερη κινητήριος δύναμη πίσω από τη συνεργατική και την αλτρουιστική συμπεριφορά. Σε πραγματικές συνθήκες τα άτομα κάνουν συγκρίσεις κόστους-οφέλους προκειμένου να επιβάλουν οποιαδήποτε μορφή τιμωρίας. Εκτός από τις παρορμητικές συναισθηματικές αντιδράσεις στις αποφάσεις τους εμπλέκονται και συνειδητοί γνωσιακοί μηχανισμοί υπολογισμού των οφελών. Ο ρόλος της άμεσης αμοιβαιότητας και των ανταποδοτικών στρατηγικών, καθώς και οι μηχανισμοί δημιουργίας φήμης παίζουν εξαιρετικά σημαντικό ρόλο για τη διατήρηση συνεργατικών συμπεριφορών στις ανθρώπινες κοινωνίες.

Η εφαρμογή κυρώσεων αναμφισβήτητα αποτελεί καθοριστικό παράγοντα για την ενίσχυση και την εδραίωση της συνεργασίας σε κοινωνικές ομάδες. Η ύπαρξη της αλτρουιστικής τιμωρίας μας βοηθά ιδιαίτερα να εξηγήσουμε τις περιπτώσεις εκείνες της συνεργασίας όπου οι άλλες θεωρίες αφήνουν κενά, όπως οι μη επαναλαμβανόμενες αλληλεπιδράσεις μεταξύ ξένων. Εξελικτικά μοντέλα δείχνουν ότι η εισαγωγή της σε μια κοινωνική προσομοίωση μπορεί να αποτελέσει το στοιχείο κλειδί για την σταθεροποίηση της συνεργασίας στην ανθρώπινη εξελικτική ιστορία.

Ωστόσο, η αποδοχή της αλτρουιστικής τιμωρίας ως του βασικότερου παράγοντα για την εξελικτική σταθεροποίηση της συνεργασίας χρειάζεται πολλή προσοχή. Η αλληλεπίδραση στις πραγματικές κοινωνίες περιλαμβάνει παραμέτρους που δεν αντιπροσωπεύονται επαρκώς στα

συμπεριφορικά πειράματα και τις εξελικτικές προσομοιώσεις. Η αλτρουιστική τιμωρία είναι πολύ πιθανό να έπαιξε ένα σημαντικό ρόλο στα πρώτα εξελικτικά στάδια της ανάπτυξης συνεργατικών προτύπων στις ανθρώπινες κοινωνίες. Παρόλο που το κίνητρο πίσω από τη συμπεριφορά αυτή παρέμεινε αναλλοίωτο, η εξέλιξη της ανθρώπινης κοινωνικότητας φαίνεται ότι δημιούργησε εναλλακτικά κανάλια για την επιβολή και τη διατήρηση των κοινωνικών προτύπων. Το βασικό χαρακτηριστικό των καναλιών αυτών είναι η μείωση του κόστους για την επιβολή της τιμωρίας και η προστασία από τον κίνδυνο διολίσθησης σε καταστάσεις αέναης αντεκδίκησης. Στην πορεία τα κανάλια αυτά απέκτησαν ολοένα και πιο οργανωμένη και επίσημη μορφή, οδηγώντας στη σημερινή υποδομή των μηχανισμών επιβολής των δεσμευτικών συμβάσεων, με κύριο εκφραστή το κράτος δικαίου στο οποίο οι σύγχρονες κοινωνίες έχουν παραχωρήσει το μονοπώλιο της βίας και της επίσημης αστυνόμευσης και επιβολής δικαιοσύνης.

Έτσι, στον πραγματικό κόσμο, η βάση για τη συνεργασία φαίνεται να μετακινήθηκε από την ανεξάρτητη και ασυντόνιστη τιμωρία σε πιο συντονισμένες μορφές επιβολής ποινών από κοινωνικές συμμαχίες. Οι κοινωνίες σχημάτισαν πιο κεντρικούς και οργανωμένους θεσμούς για την επιβολή κυρώσεων σε παραβιάσεις των κοινωνικών κανόνων, στους οποίους παραχώρησαν τα δικαιώματα τιμωρίας. Οι θεσμοί αυτοί έχουν μικρό κόστος και εμπόδια διατήρησης λόγω της νομιμοποίησης που τους χαρίζει η κοινωνική συναίνεση. Το κίνητρο πίσω από τη δημιουργία τους είναι το ίδιο συναίσθημα που προκαλεί την επιθυμία για την επιβολή της αλτρουιστικής τιμωρίας όταν δεν υπάρχει συνεννόηση και κεντρική οργάνωση και η ανάληψη προσωπικού κόστους για τιμωρία είναι το μόνο μέσο για τη νουθεσία των παραβατών και επαναφορά της δικαιοσύνης. Έτσι, στην καθημερινότητα των σύγχρονων κοινωνιών η αλτρουιστική τιμωρία περιορίζεται σε καταστάσεις όπου το κόστος της είναι πολύ μικρό και σχεδόν συμβολικό. Σε όλες τις υπόλοιπες περιπτώσεις, το στοιχείο της αλτρουιστικής τιμωρίας είναι δύσκολο να διαχωριστεί από την τιμωρία που μπορεί να έχει ένα άμεσο ή έμμεσο μελλοντικό όφελος για το άτομο.

Εν τέλει λοιπόν θα μπορούσαμε να πούμε ότι η συνεργασία στις ανθρώπινες κοινωνίες βασίζεται τόσο σε εγωιστικά όσο και σε αλτρουιστικά από υλική σκοπιά κίνητρα, η ύπαρξη των οποίων βασίζεται σε γνωσιακούς μηχανισμούς στάθμισης κόστους και οφέλους, σε νευρωνικά κυκλώματα επεξεργασίας ανταμοιβών και σε παρορμητικές συναισθηματικές αντιδράσεις . Τα κίνητρα αυτά αποτελούν προϊόν της μακραίωνης εξελικτικής διαδικασίας και συνυπάρχουν στα άτομα, επηρεάζοντας τις επιλογές τους. Έτσι εντός των κοινωνιών διαμορφώνονται δυναμικές

σχέσεις αλληλεπιδράσεις μεταξύ των ατόμων όπου οι επιλογές διαπλέκονται και αλληλοεπηρεάζονται, διαμορφώνοντας μια διαρκώς ταλαντευόμενη ισορροπία μεταξύ εγωιστικών και αλτρουιστικών συμπεριφορών. Η ισορροπία αυτή μεταβάλλεται διαρκώς και μπορεί να κινείται από το έντονα εγωιστικά μέχρι έντονα αλτρουιστικά άκρα. Το σημείο στο οποίο αυτή βρίσκεται κάθε στιγμή καθορίζει και τον αντίστοιχο βαθμό συνεργασίας στο κοινωνικό σύνολο.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Andreoni, J. & Miller, J. (2002), "Giving according to Garp: an experimental test of the consistency of preferences for altruism." *Econometrica* 70: 737-753.

Andreoni, J., Erard, B. & Feinstein, J. (1998), "Tax compliance." *Journal of Economic Literature* 36: 818-860.

Axelrod, R. (1984), "*The evolution of cooperation*." Penguin.

Axelrod, R., W. D. Hamilton (1981), "The Evolution of Cooperation." *Science* 211: 1390-1396.

Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., & Sugden, R. (2009). "*Experimental economics: Rethinking the rules*." Princeton University Press.

Batson, D. C. (1991), "*The Altruism Question*." Lawrence Erlbaum Associates, Hillsdale, NJ.

Bechara, A. Damasio, H. & Damasio, A.R. (2000), "Emotion, Decision making and the orbitofrontal cortex" *Cereb. Cortex* 10: 295-307.

Bendor, J. & Swistak, P. (2001), "The evolution of norms." *Am. J. Sociol*. 106: 1493-1545.

Berg, J., Dickhaut, J. & McCabe, K. (1995), "Trust, reciprocity, and social history." *Games and Economic Behaviour* 10: 122-142.

Binmore, K. (1999), "Why experiment in economics?" *Economic Journal* 109: F16–24.

Binmore, K. (2006), "Why do people cooperate?" *Politics, Philosophy and Economics* 5: 81-96.

Bolton, G.E. & Ockenfels, A. (2000), "A Theory of Equity, Reciprocity and Competition." *American Economic Review* 100: 166-193

Bowles, S., Choi, J.-K. & Hopfensitz, A. (2003), "The co-evolution of individual behaviours and social institutions." *J. Theor. Biol*. Jul 21;223(2): 135-47.

Bowles, S. & Gintis, H. (2004), "The evolution of strong reciprocity: Cooperation in heterogeneous populations." *Theoretical Population Biology* 65: 17-28.

Boyd, B. & Richerson P.J. (1988), "The Evolution of Reciprocity in Sizable Groups." *Journal of Theoretical Biology* 132(3): 337-56.

Boyd, R. & Richerson, P. (1992), "Punishment allows the evolution of cooperation (or anything else) in sizable groups." *Ethology and Sociobiology* 13: 171-195.

Boyd, R., Bowles, S. & Gintis, H. (2010), "Coordinated punishment of defectors sustains cooperationand can proliferate when rare." *Science* 328: 617-620.

Boyd, R., Gintis, H., Bowles, S. & Richerson, P. (2003), "The evolution of altruistic

punishment." *Proceedings of the National Academy of Sciences* 100: 3531-3535.

Boyd, R., Richerson, P. J. (2005), "*The origin and evolution of cultures*." Oxford University Press.

Bryan, J.H. (1971), "Model Affect and Children's Imitative Altruism." *Child Development* 42(6): 2061–2065.

Burnham, T.C. & Johnson, D.P. (2005), "The biological and evolutionary logic of human cooperation." *Analyse & Kritik* 27: 113-135.

Camerer, C.F. & Thaler, R.H. (1995), "Ultimatums, Dictators and Manners." *Journal of Economic Perspectives* 9: 209-19.

Camerer, C.F. (2003), "Behavioral game theory." Princeton: Princeton University Press.

Cameron, L. A. (1999), "Raising the stakes in the ultimatum game: Experimental evidence from Indonesia." *Econ. Inq*. 37: 47-59.

Carpenter, J.P., Daniere, A.G. & Takahashi, L.M. (2004), "Cooperation, trust, and social capital in Southeast Asian urban slums." *Journal of Economic Behavior and Organization* 55: 533–551.

Casari, M. & Luini, L. (2009), "Cooperation under alternative punishment institutions: An experiment." *Journal of Economic Behavior and Organization* 71: 273-282.

Casari, M. (2007), "Emergence of endogenous legal institutions: property rights and community governance in the Italian Alps." *Journal of Economic History* 67: 191-226.

Charness, G. & Rabin, M. (2000), "Social Preferences: Some Simple Tests and a New Model." Mimeo, University of California at Berkeley. http://escholarship.org/uc/item/46j0d6hb

Dawes, C.T., Fowler, J.H., Johnson, T., McElreath, R. & Smirnov, O. (2007), "Egalitarian motives in humans.'' *Nature*, Vol 446 : 794-796

Dawes, R. M. (1980), "Social dilemmas." Annu. Rev. Psychol. 31: 169-193.

de Quervain, D.J.F., Fischbacher U., Treyer, V., Schelhammer, M., Schnyder, U., Buck, A., Fehr, E. (2004), "The Neural Basis of Altruistic Punishment." *Science*, vol 305: 1254-1258.

Delgado, M.R., Stenger, V.A. & Fiez, J.A. (2004), "Motivation-dependent Responses in the Human Caudate Nucleus." *Cereb. Cortex* 14 (9): 1022-1030.

Dreber, A., Rand, D.G., Fudenberg, D. & Nowak, M.A. (2008), "Winners don't punish." Nature 452: 348-351.

Dufwenberg, M. & Kirchsteiger, G. (1998), "A Theory of Sequential Reciprocity." Discussion

Paper, CentER, Tilburg University.
http://econpapers.repec.org/paper/dgrkubcen/199837.htm

Egas, M. & Riedl, A. (2008), "The economics of altruistic punishment and the maintenance of cooperation." *Proceedings of the Royal Society* B 275: 871-878.

Falk, A. & Fischbacher, U. (2006), "A Theory of Reciprocity." *Games and Economic Behavior*, 54(2): 293-315.

Fehr, E. & Fischbacher, U. (2004), "Third-party punishment and social norms." *Evolution and Human Behavior* 25: 63-87.

Fehr, E. & Gächter, S. (2000), "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90: 980-994.

Fehr, E. & Schmidt, K.M. (1999), "A Theory of Fairness, Competition and Co-operation." *Quarterly Journal of Economics* 114: 817-868.

Fehr, E., Fischbacher, U. & Gächter, S. (2002), "Strong reciprocity, human cooperation, and the enforcement of social norms." *Hum. Nat*. 13: 1-25.

Fehr, E., Fischbacher, U. (2003), "The nature of human altruism." *Nature* 425: 785 – 791.

Fehr, E., Kirchsteiger, G. & Riedl, A. (1993), "Does fairness prevent market clearing? An experimental investigation." Quarterly Journal of Economics 108: 437-460.

Fessler, D. (2002), "Windfall and socially distributed willpower: The psychocultural dynamics of rotating Savings and Credit Associations in a Bengkulu Village." *Ethos*, 30: 25–48.

Fischbacher, U., Gächter, S., & Fehr, E. (2001), "Are people conditionally cooperative? Evidence from a public goods experiment." *Economics Letters* 71: 397-404.

Fong, C. (2001), "Social preferences, self-interest, and the demand for redistribution." *Journal of Public Economics* 82: 225-246.

Forsythe, R., Horowitz, J. L., Savin, N. E. & Sefton, M. (1994), "Fairness in simple bargaining experiments." *Game Econ. Behav*. 6: 347-369.

Fowler, J.H. (2005), "Altruistic punishment and the origin of cooperation." *Proceedings of the National Academy of Sciences,* vol. 102, no. 19: 7047–7049.

Gintis, H. (2000), "Strong reciprocity and human sociality." *J. Theor. Biol*. 206: 169-179.

Gintis, H. (2003), "The hitchhiker's guide to altruism: Gene-culture co-evolution and the internalization of norms." *J. Theor. Biol*. 220: 407-418.

Gintis, H., (2009), "*Bounds of reason. Game theory and the unification of the behavioral*

sciences." Princeton University Press.

Gintis, H., Boyd, R., Bowles, S. & Fehr, E. (2005), "*Moral sentiments and material interests: The foundations of cooperation in economic life*." MIT Press.

Gintis, H., Smith, E.A., Bowles, S. (2001), "Costly signaling and Cooperation." *J. theor. Biol.* 213: 103-119.

Glimcher, P. W., Camerer C. F., Fehr E., Poldrack R. (2009), In "*Neuroeconomics. Decision making and the brain."* (eds Glimcher, P. W., Camerer C. F., Fehr E., Poldrack R.). Chapter 1: 1-12. Elsevier Inc.

Grusec, J. E. (1971), "Power and the Internalization of Self-Denial." *Child Development* 42(1): 93–105.

Guala, F. (2005), "*The methodology of experimental economics*." Cambridge University Press.

Guala, F. (2012), "Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate." *Behav Brain Sci*.35(1):1-15

Gürerk, O., Irlenbusch, B. & Rockenbach, B. (2006), "The competitive advantage of sanctioning institutions." *Science* 312: 108-111.

Güth, W., Schmittberger, R. & Schwarze B. (1982), "An Experimental Analysis of Ultimatium Bargaining." *Journal of Economic Behavior and Organization* 3: 367-88.

Hamilton, W.D. (1964), "Genetical Evolution of Social Behavior I, II." *Journal of Theoretical Biology* 7(1): 1-52.

Harris, J. R. (1998), "*The Nurture Assumption:Why Children turn out the way they do*." New York: Touchstone.

Hart, O., Moore, J. (1998), "Foundations of incomplete contracts." Working paper 6726. http://www.nber.org/papers/w6726.

Hayashi, N., Ostrom, E., Walker, J. & Yamagishi, T. (1999), "Reciprocity, trust, and the sense of control--a cross-societal study." *Rational. Soc*. 11: 27-46.

Henrich, J. & Boyd, R. (1998), "The evolution of conformist transmission and the emergence of between-group differences." *Evol. Hum. Behav*. 19: 215-242.

Henrich, J. & Boyd, R. (2001), "Why people punish defectors-weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas." *J. Theor. Biol*. 208: 79-89.

Henrich, J. & Henrich, N. (2007), "*Why humans cooperate: A cultural and evolutionary*

explanation." Oxford University Press.

Henrich, J. et al. (2001), "In search of Homo economicus: behavioral experiments in 15 small-scale societies." *Am. Econ. Rev*. 91: 73-78.

Henrich, J. et al. (2006), "Costly punishment across human societies." *Science* 312: 1767–1770.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E. & Gintis, H. (2004), "*Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*." Oxford University Press.

Henrich,J. (2000), "Does culture matter in economic behavior? Ultimatum Game Bargaining among the Machiguenga of the Peruvian Amazon." *American Economic Review* 90 (4): 973-979.

Hill, K. (2002), "Altruistic cooperation during foraging by the Ache, and the evolved human predisposition to cooperate." *Hum. Nat*. 13: 105-128.

Hoffman, E., McCabe, K., Shachat, K. & Smith, V. (1994), "Preferences, property rights and anonymity in bargaining games." *Game Econ. Behav*. 7: 346-380.

Isaac, R. M. & Walker, J. M. (1988), "Group-size effects in public-goods provision--the voluntary contributions mechanism." *Q. J. Econ*. 103: 179-199.

Kahneman, D., Knetsch, J.L. & Thaler, R.H. (1991), "Anomalies: The endowment effect, loss aversion, and status quo bias." Journal of Economic Perspectives 5: 193-206.

Kiyonari, T., Tanida, S. & Yamagishi, T. (2000), "Social exchange and reciprocity: confusion or a heuristic." *Evol. Hum. Behav*. 21: 411-427.

Knutson, B., Westdorp,A., Kaiser, E. & Hommer, D. (2000), "FMRI visualization of brain activity during a monetary incentive delay task." Neuroimage 12: 20-27.

Ledyard, J. (1995), In "*Handbook of Experimental Economics"* (eds Kagel, J. & Roth, A.): 111-194. Princeton Univ. Press.

Levine, D. K. (1998), "Modeling altruism and spitefulness in experiments." *Rev. Econ. Dynam*. 1: 593-622.

McAndrew, F. (2002), "New Evolutionary Perspectives on Altruism - Multilevel-Selection and Costly-Signaling Theories." *Current directions in Psychological Science* 11 (Issue 2): 79-82

Messick, D. & Brewer, M. (1983), In "*Review of Personality and Social Psychology"* (ed. Wheeler, L.).Sage Publ., Beverly Hills.

Nikiforakis, N. & Engelmann, D. (2010), "Altruistic punishment and the threat of feuds." Department of Economics, University of Melbourne.

Nikiforakis, N. (2008), "Punishment and counter-punishment in public good games: Can we really govern ourselves?" *Journal of Public Economics* 92: 91-112.

Nowak, M. & Highfield, R. (2011), "*Supercooperators*." Free Press.

Nowak, M. & Sigmund, K. (1998), "Evolution of Indirect Reciprocity by Image Scoring." *Nature* 393: 573-577.

O'Doherty, J., Dayan, P., Schultz, J., et al., (2004), "Dissociable roles of ventral and dorsal striatum in instrumental conditioning." *Science* 304: 452-454.

Ostrom, E. (1990), "*Governing the commons: The evolution of institutions for collective action."* Cambridge University Press.

Ostrom, E. (2000), "Collective action and the evolution of social norms." *Journal of Economic Perspectives* 14: 137-158.

Ostrom, E., Gardner, R., & Walker J. (1994), "*Rules, games and common-pool resources*." The University of Michigan Press.

Ostrom, E., Walker, J. & Gardner, R. (1992), "Covenants with and without a sword: self-governance is possible." *American Political Science Review* 86: 404-417.

Rabin, M. (1993), "Incorporating fairness into game theory and economics." *Am. Econ. Rev*. 83: 1281-1302.

Ramnani, N. & Owen, A.M. (2004), "Anterior prefrontal cortex: insights into function from anatomy and neuroimaging." *Nature Reviews Neuroscience* 5:184-194.

Rilling, J. K. et al. (2002), "A neural basis for social cooperation." *Neuron* 35, 395-405.

Richerson, P.J., Boyd, R.T. & Henrich,J. (2003), In "*Genetic and Cultural Evolution of Cooperation."* (eds Hammerstein, P.). Chapter 19: 357-388. MIT Press.

Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M. & Zamir S. (1991), "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study." *American Economic Review* 81: 1068-95.

Sachs, J. L., Mueller, U. G., Wilcox, T. P. & Bull, J. J. (2004), "The evolution of cooperation." *Q. Rev. Biol*. 79: 135–160.

Sanfey, A.G., Rilling, J.K., Aaronson, J.A., Nystrom, L.E., & Cohen, J.D. (2003), "The neural basis of economic decision-making in the ultimatum game." *Science*, 300: 1755–58.

Schultz W. & Romo, R. (1988), “Neuronal activity in the monkey striatum during the initiation of movements.” *Experimental brain research.* Volume 71, Number 2 (1988): 431-436.

Sigmund, K., Hauert, C. & Nowak, M. A. (2001), “Reward and punishment.” *Proceedings of the National Academy of Sciences* USA 98: 10 757–10 762.

Silk, J. (2009). In “*Neuroeconomics. Decision making and the brain*.” (eds Glimcher, P. W., Camerer C. F., Fehr E., Poldrack R.). Chapter 18: 269-284. Elsevier Inc.

Simon, H. (1990), “A mechanism for social selection and successful altruism.” *Science* 250: 1665–1668.

Singer, T. (2009), In “Neuroeconomics. *Decision making and the brain*.” (eds Glimcher, P. W., Camerer C. F., Fehr E., Poldrack R.). Chapter 17: 251-268. Elsevier Inc.

Slonim, R. & Roth A.E. (1998), “Financial Incentives and Learning in Ultimatum and Market Games: An Experiment in the Slovak Republic.” *Econometrica* 65: 569-596.

Sober, E., & Wilson, D. S. (1998), “*Unto others—The evolution and psychology of unselfish behavior*.” Cambridge, MA: Harvard University Press.

Starmer, C. (1999), “Experiments in economics … (should we trust the dismal scientists in white coats?)” *Journal of Economic Methodology* 6: 1-30.

Takahasi, k. (1999), “Theoretical aspects of the mode of transmission in cultural inheritance.” *Theor. Popul. Biol*. 55: 208-225.

Trivers, R. L. (1971), “The Evolution of Reciprocal Altruism.” *Quarterly Review of Biology* 46: 35-57.

Trivers, R.L. (2004), “Behavioural evolution: Mutual benefits at all levels of life.” *Science* 304: 964-965.

Ule, A., Schram, A., Riedl, A., & Cason, T.N. (2009), “Indirect punishment and generosity toward strangers.” *Science* 326: 1701-1704.

West, S. A., Pen, I. & Griffin, A. S. (2002), “Cooperation and competition between relatives.” *Science* 296: 72–75.

West, S.A.; Griffin, A.S. & Gardner, A. (2007), "Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection". *Journal of Evolutionary Biology* 20 (2): 415–32.

Wilkinson, G. S. (1984), “Reciprocal food sharing in the vampire bats.” *Nature* 308: 181–84.

Xiao, E. & Houser, D. (2005), “Emotion expression in human punishment behaviour.”

Proceedings of the National Academy of Science 102: 7398-7401.

Yamagishi, T. (1986), “The provision of a sanctioning system as a public good.” *Journal of Personality and Social Psychology* 51: 110-116.