

PHILOSOPHICAL READINGS

ONLINE JOURNAL OF PHILOSOPHY

RR
VI

NUMBER 3
AUTUMN 2014

Special Issue on:
**EXPERIMENTING ECONOMICS:
USING BEHAVIOURAL
SCIENCES IN RESEARCH**

Guest Editor:
Luciano Canova

ARTICLES

- Luciano Canova*
Foreword 3
- Tommaso Reggiani*
Una breve nota sull'etica comportamentale 5
- Carlo Canepa – Matteo Motterlini*
La rivoluzione "gentile" delle politiche basate sull'evidenza.
Considerazioni epistemologiche 9
- Alessandra Smerilli*
Theories of Team Reasoning 24
- Daniela Grieco*
On the Demand and Effectiveness of Reward in Voluntary
Contribution Mechanisms. A Note. 35

REVIEWS 46

Maria Carolina Foi, *La giurisdizione delle scene. I drammi politici di Schiller* (Macerata: Quodlibet, 2013) (L.A. Macor); Barbara Scapolo, *Leggere Timore e Tremore di Kierkegaard* (Como – Pavia: Ibis, 2013) (A. Granito).

ABSTRACTS AND INDEXING 54

ISSN 2036-4989



Philosophical Readings

Online Journal of Philosophy

Philosophical Readings, ISSN 2036-4989, features articles, discussions, translations, reviews, and bibliographical information on all philosophical disciplines. *Philosophical Readings* is devoted to the promotion of competent and definitive contributions to philosophical knowledge. Not associated with any school or group, not the organ of any association or institution, it is interested in persistent and resolute inquiries into root questions, regardless of the writer's affiliation. The journal welcomes also works that fall into various disciplines: religion, history, literature, law, political science, computer science, economics, and empirical sciences that deal with philosophical problems. *Philosophical Readings* uses a policy of blind review by at least two consultants to evaluate articles accepted for serious consideration. *Philosophical Readings* promotes special issues on particular topics of special relevance in the philosophical debates. *Philosophical Readings* occasionally has opportunities for Guest Editors for special issues of the journal. Anyone who has an idea for a special issue and would like that idea to be considered, should contact the Executive editor.

Executive editor: Marco Sgarbi, *Università Ca' Foscari - Venezia*.

Associate editor: Eva Del Soldato, *University of Pennsylvania*.

Assistant editor: Valerio Rocco Lozano, *Universidad Autónoma de Madrid*.

Review editor: Laura Anna Macor, *Università di Padova*.

Editorial Advisory Board: Laura Boella, *Università Statale di Milano*; Elio Franzini, *Università Statale di Milano*; Alessandro Ghisalberti, *Università*

Cattolica di Milano; Piergiorgio Grassi, *Università di Urbino*; Margarita Kranz, *Freie Universität Berlin*; Sandro Mancini, *Università di Palermo*; Massimo Marassi, *Università Cattolica di Milano*; Roberto Mor-dacci, *Università Vita e Salute San Raffaele di Milano*; Ugo Perone, *Università del Piemonte Orientale*; Riccardo Pozzo, *Consiglio Nazionale delle Ricerche*; José Manuel Sevilla Fernández, *Universidad de Sevilla*.

Editorial Board: Raphael Ebgi, *Università San Raffaele di Milano*; Luca Gili, *Katholieke Universiteit Leuven*; Pierpaolo Marrone, *Università di Trieste*; Eugenio Refini, *The Johns Hopkins University*; Alberto Vanzo, *University of Warwick*; Francesco Verde, *Università "La Sapienza" di Roma*; Antonio Vernacotola, *Università di Padova*.

Board of Consultants: This board has as its primary responsibility the evaluation of articles submitted for publication in *Philosophical Readings*. Its membership includes a large group of scholars representing a variety of research areas and philosophical approaches. From time to time, *Philosophical Readings* acknowledges their service by publishing the names of those who have read and evaluated manuscripts in recent years.

Submissions: Submissions should be made to the Editors. An abstract of not more than seventy words should accompany the submission. Since *Philosophical Readings* has adopted a policy of blind review, information identify the author should only appear on a separate page. Most reviews are invited. However, colleagues wishing to write a review should contact the Executive editor. Books to be reviewed, should be sent to the Executive editor.

Introduzione

L'economia è stata definita da Thomas Carlyle *la scienza triste* e, complice la più grave crisi dal secondo Dopoguerra ad oggi, dentro la quale siamo ancora immersi, è una disciplina che sta attraversando una profonda trasformazione, sottoposta agli strali di chi considera il pensiero standard come inscalfibile a qualunque forma di innovazione e bisognoso di un cambio di passo.

In realtà, senza dover scomodare la struttura delle rivoluzioni scientifiche di Kuhn, a livello epistemologico l'economia è disciplina viva e vegeta, con una profonda multidisciplinarietà e un costante dialogo tra approcci differenti. Questo numero di *Philosophical Readings* raccoglie quattro contributi che, in modo diverso, fanno riferimento all'utilizzo del metodo sperimentale per l'analisi del comportamento economico. Si tratta di una letteratura di per sé non nuovissima, dato che Vernon Smith e Daniel Kahneman (insigniti del Premio Nobel nel 2002) hanno pubblicato i primi seminali contributi sul finire degli anni '70 del secolo scorso. E si tratta di un approccio ormai riconosciuto all'interno della comunità accademica come in grado di gettare una luce più approfondita sulle dinamiche del comportamento individuale, con particolare riferimento proprio alle scelte economiche.

Il primo saggio, di Tommaso Reggiani (Università di Colonia), è una breve nota di approfondimento sull'etica comportamentale e sugli interessanti contributi di tale approc-

cio con riferimento alla soluzione di dilemmi etici quali il trolley problem. Il focus circa il comportamento effettivo di persone esposte a dilemmi morali è la nuova nicchia di studio ove si collocano i *behavioral ethicists*. Essi attingono da un retroterra filosofico per applicare quindi nozioni e metodologie di indagine tipiche della psicologia e dell'economia comportamentale / sperimentale. Questa nuova classe di studiosi mira a fotografare il comportamento reale delle persone, descrivere come i fattori legati al contesto ambientale e sociale influenzino il discernimento morale, al fine di individuare opportunità per agevolare un maggior numero di comportamenti etici tramite l'implementazione di piccoli interventi (*nudge*) da parte dei policy makers.

Strettamente connesso a questo primo intervento, il saggio di Carlo Canepa e Matteo Motterlini (Università San Raffaele) estende la riflessione all'ambito macroeconomico dei decisori pubblici entrando nel dettaglio della filosofia politica del *nudge* (la spinta gentile). La rivoluzione *nudge* mostra infatti come far leva sui fattori sociali e cognitivi che influenzano le decisioni per promuovere comportamenti virtuosi, indirizzando la libertà di scelta degli individui, senza limitarla. L'*evidence-based policy*, dunque, introduce la sperimentazione per valutare quali politiche effettivamente funzionino e quali no, sulla base dell'evidenza data dai risultati raccolti, sottraendo il disegno di *public policy* agli sterili dibattiti pregiudizievole e ideologizzati.

Il contributo di Alessandra Smerilli (Università Pontificia Salesiana) è un'estensione

del concetto di razionalità, tradizionalmente considerato in economia come frutto di azioni esclusivamente individuali. Smerilli sviluppa da un punto di vista normativo un approccio originale teso a dimostrare l'emergere di esiti cooperativi in un classico dilemma del prigioniero, con particolare attenzione alle negoziazioni in ambito ambientale. L'approccio è basato sulla nozione della *we-rationality*, sviluppata da Hollis nel 1998. I modelli *we-thinking* trattano i gruppi come se fossero individui in grado di operare scelte autonome che configurano l'emergere di un *frame* del 'noi'.

Infine, la nota di Daniela Grieco (Università Bocconi) si spinge addentro alla logica della cooperazione come dinamica emergente dall'interazione strategica di individui facenti parte di un gruppo, presentando i risultati dell'approccio sperimentale nell'analizzare esiti *pro-social* piuttosto che discriminatori e individuando alcune leve comportamentali che potrebbero favorire i primi rispetto ai secondi.

Questi brevi saggi e l'intero numero di *Philosophical Readings* sono il risultato di una selezione attenta, tesa a fornire un assaggio, uno sguardo aperto sulle potenzialità dell'approccio sperimentale nella disciplina economica. Le scienze comportamentali e la loro integrazione con la scuola cosiddetta neoclassica possono contribuire a informare la realtà e il design di politiche pubbliche sempre centrate sul cittadino e fondate su un'evidenza empirica robusta ai tempi della società *data driven*.

Luciano Canova

Una breve nota sull'etica comportamentale

Tommaso Reggiani
Universität zu Köln
(Deutschland)

1. Introduzione

Il tradizionale metodo di ricerca e di insegnamento riguardante l'ambito filosofico adotta

un approccio prettamente normativo. Esso si concentra intorno alla questione su come la persona dovrebbe comportarsi quando esposto di fronte a dilemmi etici. In questa nota descriviamo gli orientamenti etici tradizionali per poi procedere a esporre esempi di come un *approccio positivo di impronta comportamentale all'etica* possa arricchire con profitto le prospettive classiche. In questo ambito l'*etica comportamentale* si delinea come lo studio dei comportamenti che sistematicamente e prevedibilmente divergono rispetto all'intuizione comune di azione moralmente corretta. Focalizzandosi su una dimensione positiva-descrittiva, piuttosto che su un'ottica puramente normativa, l'analisi comportamentale del fenomeno etico risulta essere particolarmente interessante al fine di investigare cause e determinanti che inducono persone di buona volontà a infrangere norme etiche e morali. Le considerazioni e le raccomandazioni pratiche che scaturiscono da questo approccio di analisi risultano essere particolarmente interessanti e vastamente

applicabili a casi concreti e ricorrenti nell'ambito dell'etica degli affari.

2. Etica classica

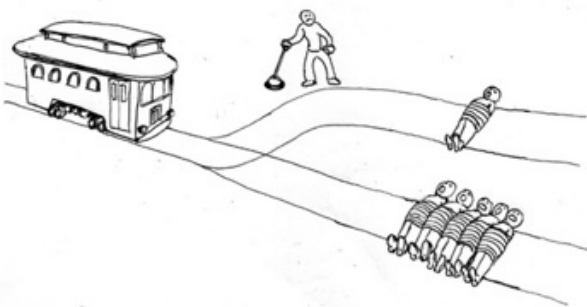
Quando si parla di corsi di formazione in ambito etico o di ricerca accademica nel medesimo contesto, la materia è stata tipicamente appannaggio di filosofi. Le nozioni impartite nelle aule hanno avuto, spesso, l'obiettivo esclusivo di aiutare gli studenti a discernere le differenti tradizioni filosofiche al fine di comprendere gli elementi costitutivi di un dilemma etico ed il conseguente comportamento da assumere. Le nozioni sono state tradizionalmente veicolate con un approccio narrativo neglignendo l'importanza di calare le nozioni all'interno di casi concreti. Con più stretto riguardo all'etica degli affari - per la maggior parte dei casi - gli studiosi hanno sposato un orizzonte prescrittivo mutuando, quindi, spunti dalla filosofia teoretica al fine di definire gli standard di comportamento..

Come risultato, la *business ethics* si è sviluppata valutando le pratiche di lavoratori, managers e organizzazioni nel loro insieme, da un punto di vista squisitamente moralistico. In questo solco metodologico la *teoria dello sviluppo cognitivo* costituisce uno dei nuclei della materia. Il giudizio etico risulta quindi essere determinato dal livello di sofisticazione morale attraverso il quale il dilemma morale viene analizzato e processato dalla persona. Seguendo questa teoria, persone che hanno raggiunto uno stadio avanzato di discernimento morale tendono ad

adottare comportamenti eticamente più adeguati rispetto a persone che hanno raggiunto uno stadio di sofisticazione morale di livello inferiore.

Sebbene una persona possa avere raggiunto un alto livello di sofisticazione morale, essa in alcune circostanze potrebbe arrestare la sua valutazione morale a livelli inferiori e questo determina il verificarsi di azioni immorali da parte di persone generalmente ritenute eticamente corrette. Lo studio delle teorie di sviluppo cognitivo e morale è stato sovente accompagnato dall'esposizione dei dibattiti e delle distinzioni che esistono all'interno del mondo filosofico e rimangono irrisolte da secoli. Un esempio classico di distinzione fra differenti correnti normative riguardanti il giudizio etico è rappresentata da Utilitarismo vs. Deontologia.

Tradizionalmente il "trolley problem" (forse, breve articolazione della frase e spiegazione? C'è solo l'immagine)



Da una prospettiva teoretica, i filosofi tradizionali stanno ancor oggi dibattendo circa l'opportunità di ritenere moralmente più accettabile una prospettiva utilitarista vs deontologica a questa tipologia di fattispecie. Da un punto di vista empirico, evidenza e dati consentono oggi agli studiosi interessati di dipartire da questo dibattito per dirigersi

verso lo studio delle condizioni a seconda delle quali la gente reale è più propensa ad implementare un criterio piuttosto che l'altro quando si confronta con dilemmi etici di tale tipologia.

Sebbene le dispute filosofiche possano apparire inconcludenti, i filosofi classici - anche in linea con le conclusioni raggiunte dall'approccio dello sviluppo cognitivo sopra citato - , argomentano come la pratica sistematica della dialettica e del confronto filosofico sia un punto centrale per lo sviluppo morale a livello individuale. Se si accetta questa prospettiva, allora essa implicherebbe direttamente la promozione di un'attitudine alla cittadinanza più responsabile e una decisa avversione alle ingiustizie. Tuttavia un filosofo (chi?), molto interessato alle questioni metodologiche e sperimentali, ha fatto notare come l'evidenza empirica circa questa convinzione si diriga in opposta direzione. Analizzando i dati di molteplici biblioteche accademiche, egli mostra come professori e ricercatori universitari attivi nel campo della filosofia sono molto più propensi a incorrere in gravi ritardi nelle riconsegne dei volumi (quando non addirittura furti!) rispetto accademici impegnati negli altri campi di sapere.

3. *Evidence based ethics*: un nuovo approccio

Fatti come questo appaiono di difficile comprensione se si adottano lenti normative tipiche dalla filosofia tradizionale che si è concentrata essenzialmente sul *come le persone si dovrebbero comportare*.

Con questo non si vuole sminuire la ricchezza di sapere generatasi attraverso questo approccio. Tuttavia esso ha contribuito innegabilmente a prevenire ogni interesse per l'evidenza empirica circa questi fenomeni e conseguentemente l'individuazione di strumenti per migliorare le scelte e i comportamenti eticamente rilevanti delle persone. Il focus circa il comportamento effettivo di persone esposte a dilemmi morali è la nuova nicchia di studio ove si collocano i *behavioral ethicists*.

Essi attingono da un retroterra filosofico per applicare quindi nozioni e metodologie di indagine tipiche della psicologia e dell'economia comportamentale / sperimentale. Questa nuova classe di studiosi mira a fotografare il comportamento reale delle persone, descrivere come i fattori legati al contesto ambientale e sociale influenzino il discernimento morale, al fine di individuare opportunità per agevolare un maggior numero di comportamenti etici tramite l'implementazione di piccoli interventi (*nudge*) da parte dei policy makers. Come primo passo di questo programma di ricerca, è quindi necessario comprendere presupposti e conseguenze tanto dei comportamenti etici quanto delle azioni immorali. Conseguentemente, come secondo passo, sarà necessario identificare strumenti concreti sia a livello individuale che a livello di disegno delle istituzioni per poter guidare attitudini immorali in persone che normalmente non vorrebbero in alcun modo infrangere le convenzioni etico-sociali.

4. Esempi e applicazioni

Il primo step dell'agenda di ricerca è già avviato e i primi risultati appaiono molto interessanti.

Per esempio è emerso come le persone abbiano in mente un bilancio della propria moralità (*moral balancing*). Questo significa che, fissato uno standard di moralità a livello individuale e su un determinato arco temporale, l'aver agito al di sopra di tale soglia ci induce (ci legittima in un certo senso...) a compiere qualche comportamento immorale conseguente. Questo vale anche in senso contrario: infatti, se oggi abbiamo agito in modo non etico, è molto probabile che domani - quando fronteggeremo un nuovo dilemma morale - al fine di tenere bilanciato il nostro conto di moralità ci sentiremo obbligati ad agire seguendo canoni morali in modo stretto.

La legittimazione da parte di una autorità costituisce un ulteriore aspetto per giustificare comportamenti immorali. In un esperimento di tipo sociologico, studenti dell'esclusiva università di Stanford furono assegnati per due settimane al ruolo di guardia o prigioniero al fine di animare un gioco di simulazione (inserirre reference). L'esperimento fu interrotto solo dopo pochi giorni a causa delle crudeltà inflitte dalle (finte) guardie ai (finti) carcerati. In un contesto normale nessuno studente di Stanford avrebbe mai inflitto tali crudeltà ai propri colleghi di studi. Questo studio mette in evidenza come fattori di contesto, quali l'assegnazione o il riconoscimento di autorità, siano in grado di influenzare pesantemente

te i giudizi etici delle persone mostrando come la morale sia malleabile e dinamica di fronte alle situazioni che fungono da scenario. Un'altra serie di esperimenti (inserire references) ha mostrato come, in qualità di spettatori di un atto immorale, siamo più propensi a ritenerlo moralmente accettabile quando il suo risultato risulta essere positivo piuttosto che negativo. Di converso, in qualità di attori siamo più indulgenti nell'adottare un comportamento immorale al fine di evitare una perdita piuttosto che nel caso di realizzare un guadagno di pari entità. In un contesto simile, le persone sono più propense a rubare quando anche i membri del proprio gruppo di riferimento (e.g. board of directors) ne traggono un beneficio piuttosto che nel caso in cui l'atto immorale porti beneficio limitatamente a chi commette l'atto moralmente repressibile.

5. Conclusione

Solo comprendendo in profondità come le persone agiscono in modo immorale, sia intenzionalmente che non-intenzionalmente, si potrà successivamente intervenire sul disegno di organizzazioni ed istituzioni al fine di promuovere comportamenti eticamente condivisi. In seconda istanza, piuttosto che prescrivere come ci si dovrebbe comportare nei confronti dei dilemmi morali o informare passivamente circa le raccomandazioni dei diversi classici della filosofia, la giovane esperienza della *behavioral ethics* ci suggerisce l'opportunità di aiutare studenti, *practitioners* e l'opinione

pubblica in generale, a capire meglio ed in modo obiettivo il proprio comportamento di fronte a scenari eticamente rilevanti. Ciò permetterà una più lucida e serena comparazione del proprio agire concreto rispetto al *benchmark* etico che si ritiene migliore e più appropriato. Solamente analizzando e prendendo coscienza delle incongruenze fra desiderio di moralità ideale e comportamenti effettivi calati nella realtà ordinaria, si potrà conseguire un eventuale scarto nella qualità dell'etica pubblica presso le diverse comunità.

La rivoluzione “gentile” delle politiche basate sull’evidenza. Considerazioni epistemologiche

.....
 Carlo Canepa – Matteo Motterlini
 Cresa, Università San Raffaele
 (Italia)

1. Introduzione

La progettazione di politiche economiche e pubbliche tradizionalmente ha tralasciato due importanti fattori. Da un lato, si è fatto scarso utilizzo dei risultati ottenuti dalle scienze comportamentali e sociali (Shafir, 2012). Dall’altro, non si è riusciti a sfruttare la forza del metodo sperimentale. La prassi è stata piuttosto quella di seguire la linea tracciata dalla *teoria economica neoclassica*, che vede ogni individuo come un razionale calcolatore della propria utilità attesa. Questo modello astratto, che alla luce dei recenti risultati sperimentali mal si adatta ai reali meccanismi di scelta degli esseri umani, ha portato, nel *policy making*, a regolare i comportamenti dei cittadini per lo più tramite la modifica di incentivi economici, l’utilizzo di divieti e di regolamentazioni.

A partire dagli ultimi cinque anni la fertile combinazione di due approcci innovativi tenta di trasformare questo panorama. La rivoluzione *nudge* mostra come far leva sui fattori sociali e cognitivi che influenzano le decisioni per promuovere comportamenti virtuosi, indirizzando la libertà di scelta degli

individui, senza limitarla. L’*evidence-based policy* introduce la sperimentazione per valutare quali politiche effettivamente funzionino e quali no, sulla base dell’evidenza data dai risultati raccolti, sottraendo il disegno di *public policy* agli sterili dibattiti pregiudizievole e ideologizzati.

2. Architettura delle scelte per un mondo più semplice

L’*architettura delle scelte* è il modo in cui sono presentate le opzioni di scelta in un processo decisionale. Come la struttura di un edificio pone dei vincoli fisici alla possibilità di muoversi e interagire con esso, così il modo in cui è strutturato lo spazio delle scelte influisce sul risultato della decisione. Ogni dettaglio può risultare importante e la capacità di condizionamento è ubiqua e mai neutrale. Nel caso della disposizione dei cibi in una mensa, per esempio, la modalità in cui questi ci vengono presentati, la grandezza dei piatti e dei bicchieri, condizionano cosa si decide di mangiare e in quali quantità. Si è potuto verificare che posizionare pietanze salutari in maniera visivamente saliente ne aumenta i consumi (Rozin et al., 2011) e ridurre il diametro delle portate diminuisce la quantità di cibo che finisce nella spazzatura (Kalbekken, Saelen, 2013)¹.

¹ Nel primo studio, all’interno di una mensa self-service, si è agito su due fattori: a) si è modificato la vicinanza delle pietanze ad alto contenuto calorico, allontanandoli di circa 25 cm dall’usuale posizione

In generale, non prendiamo mai decisioni nel vuoto, ma sempre in un dato contesto. Strutturare quest'ultimo è il compito di ogni *architetto delle scelte*. Nei processi di *policy making*, le istituzioni hanno la possibilità di far leva sui meccanismi cognitivi - sempre meglio compresi dalle neuroscienze delle decisioni e sempre più applicati dall'economia cognitiva (o comportamentale) - per orientare i cittadini verso comportamenti "virtuosi", con una "spinta gentile", a vantaggio di se stessi e della collettività.

Affinché un intervento possa definirsi un *nudge* (cf. Thaler, Sunstein, 2008), è necessario che esso:

- (i) sia un elemento di un'architettura delle scelte,
- (ii) alteri in maniera sistematica (e quindi prevedibile) il *comportamento* degli individui (componente paternalistica),
- (iii) non limiti le possibilità di scelta di quest'ultimi (componente libertaria),

del consumatore; b) si è sostituito il cucchiaino con le pinze da cucina, come strumento per servirsi. Entrambe le modifiche, che agiscono sulla localizzazione e sull'accessibilità delle portate (e non sulla diretta quantità), hanno condotto ad un calo dei consumi non salutari tra l'8% e il 16%. Nel secondo studio, all'interno di un buffet, si è agito sul diametro dei piatti. Questa variabile comunica al fruitore la quantità di cibo della quale si ritiene idonei servirsi. Diminuendola, si ottiene una riduzione dello spreco di cibo intorno al 19.5%.

- (iv) né modifichi in maniera significativa gli incentivi economici².

2.1 Opzioni di default e confronto sociale

Le *opzioni di default* rispettano queste quattro condizioni e sono un esempio di *nudge* che ha dato buona prova di sé. Esse sono le opzioni di scelta che vengono preselezionate per chi deve compiere una decisione, nel caso il soggetto non manifesti una scelta attiva differente. Un software, se non viene data indicazione diversa, è di norma installato seguendo le opzioni automatiche, non personalizzate, decise dal

² Ampliando la condizione (iv), Hausman e Welch (2010) affermano che "i nudge sono modi di influenzare le scelte senza limitarne il set o senza rendere le alternative più appetibili in termini di tempo, fatica e sanzioni sociali", quindi non solo in termini economici. Molta ambiguità persiste su questo tema nella letteratura sulle "spinte gentili", poiché queste sono per definizione restie ad incorporare l'impiego di incentivi monetari. In realtà, numerose ricerche hanno indagato i meccanismi psicologici che influenzano la ricezione di quest'ultimi. Pagare troppo, o troppo poco, per promuovere comportamenti virtuosi, può essere controproducente (Kamenica, 2012). La proposta più promettente è quella di ampliare la sfera d'azione degli approcci *nudge*. Strutturare gli ambienti di scelta in maniera cognitivamente informata può prevedere l'utilizzo di incentivi provenienti dall'esterno, se l'evidenza di successo dell'intervento è giustificata alla prova dei fatti (si veda a proposito l'efficacia dell'utilizzo di premi in lotterie nel contrastare fenomeni come l'obesità e il fumo, in Volpp et al. 2008; 2009).

produttore. Così un buon architetto delle scelte può agire facendo leva sulle abitudini e sulle consuetudini dei cittadini, creando opzioni di *default* con livelli più o meno alti di personalizzazione. In questo modo, la libertà di scelta viene indirizzata a favore dell'opzione di default, senza tuttavia risultare limitata, poiché chiunque ha facoltà di scegliere diversamente.

Nella selezione di un'assicurazione sanitaria, di un piano pensionistico, di un contratto o di un mutuo, un *policy maker* può comunicare tramite la scelta di *default* che quest'ultima è quella ottimale o il punto di riferimento con il quale confrontare le altre opzioni a disposizione (Sunstein, 2012).

Negli Stati Uniti, dal 2004, il progetto *Save More Tomorrow (SMarT Program™)* sfrutta l'automaticità nell'architettura delle scelte per far fronte ai bassi tassi di iscrizione ai piani di risparmio. I suoi ingredienti essenziali seguono il "modello delle quattro A": a) *availability* ("disponibilità"): l'accesso ai piani è il più semplice possibile; b) *automatic enrollment*: i lavoratori sono automaticamente iscritti ai piani (le disiscrizioni sono in media solo del 10%); c) *automatic investment*: è presente un portafoglio di investimenti di *default*; d) *automatic escalation*: per aumentare progressivamente il tasso di contribuzione allo schema pensionistico (solitamente basso all'inizio della carriera lavorativa) lo *SMarT Program™* fa impegnare automaticamente i contribuenti a operare una serie di contributi previdenziali sincronizzati con gli aumenti di stipendio. In meno di quattro anni, nelle compagnie in cui è stato applicato il programma, i risultati mostra-

no un aumento dei tassi di risparmio dei lavoratori dal 3.5% al 13.6% (Benartzi, Thaler, 2013)³.

Raj Chetty (Harvard University), neo vincitore della medaglia John Bates Clark, premio al miglior economista americano sotto i quarant'anni, ha mostrato, sulla base di 45 milioni di osservazioni, che in Danimarca i piani pensionistici con accantonamenti automatici sono sensibilmente più efficaci di quelli che presuppongono una contribuzione volontaria. L'85 per cento della popolazione tende infatti a mantenere inercialmente il proprio status quo di non contribuzione contro i suoi stessi interessi di lungo periodo.

Un altro fattore su cui può giocare un architetto delle scelte è il *confronto sociale*. Come si sa, le persone pongono attenzione ai comportamenti dei propri pari e tendono a conformarsi. Questo aspetto può essere utilizzato per ridurre i consumi domestici. Si è stimato che si può ottenere una diminuzione del 3.5% dell'energia elettrica consumata, inserendo semplicemente nella bolletta un confronto tra i consumi di un cittadino e quelli del suo vicinato. I meno virtuosi tendono ad adeguarsi alla media di tutti gli altri consumatori. E ciò rende tale intervento nettamente più efficace (oltre che più economico) rispetto al rialzo dei prezzi o a campagne di sensibilizzazione ambientale (Allcott, 2011)⁴.

³ Un introduttivo catalogo delle opzioni di default si trova in Goldstein et al., 2008. Per una loro efficace applicazione nell'incrementazione delle donazioni d'organi si veda Johnson, Goldstein, 2003.

⁴ Dal 2007, *Opower*, leader statunitense nel settore del risparmio energetico, sfrutta questo principio

La forza delle *norme sociali*, applicata in altri settori, ha condotto a risultati rilevanti. Comunicare ai contribuenti con una lettera che nove persone su dieci nel proprio paese pagano le tasse in tempo, o rinnovano la licenza del proprio negozio nei limiti prescritti, fa aumentare la riscossione delle tasse del 15% circa e del 7% il rinnovo delle licenze, nelle tempistiche corrette, rispetto a lettere standard ed anonime (Behavioural Insights Team, 2012a; Walsh, 2012). Comunicare ai clienti di un hotel che la maggior parte degli altri clienti riutilizza i propri asciugamani, fa aumentare del 14% il loro riutilizzo (Schultz et al., 2008). Anche in questi casi, le quattro condizioni d'intervento per avere un *nudge* sono rispettate.

Strutturare l'ambiente di scelta in maniera cognitivamente informata, facendosi guidare dall'evidenza dei risultati ottenuti una volta implementati gli interventi, è alla base degli approcci comportamentali al *policy making* (Shafir, 2012).

Su questo sfondo è necessario delineare una *competenza* mirata di chi si occupa di redigere piani e architetture delle scelte. Si noti che non si tratta di farsi dettare la politica economica pubblica partendo dalla astratta e idealizzata teoria economica della scelta razionale, ma di architettare interventi a partire da come effettivamente prendiamo le no-

comportamentale e i database in tempo reale sui consumi di 15 milioni di abitazioni. In sei anni questa strategia ha portato a risparmiare 2.5 terawattora di corrente elettrica (quantità consumata annualmente da una città di circa 400 mila abitanti) (<http://opower.com>).

stre decisioni. E' la consapevolezza della nostra razionalità limitata che può aiutarci ad arginarla, ponendo come fulcro la *trasparenza* dell'intervento. Questo è il punto centrale di un'analisi epistemologica volta a una corretta pratica del *nudge*, basata su due elementi:

- i) la formazione di architetti delle scelte *competenti*, dotati di metodo e di consapevolezza metodologica;
- ii) la raccolta di *evidenza* dell'efficacia dell'intervento e, in ultima istanza, la sua trasparenza.

Le prossime sezioni affrontano questi due aspetti.

3. Metodologia e pratica del *nudge*

La geometria sottostante alla cosiddetta *Nudge Revolution* nel *policy making* è triangolare: ai vertici stanno coloro che devono compiere una scelta (es: i cittadini), i responsabili delle modifiche al contesto (es: governi e funzionari) e gli architetti della scelta, con l'accavallarsi spesso di quest'ultimi due ruoli (Figura 1).

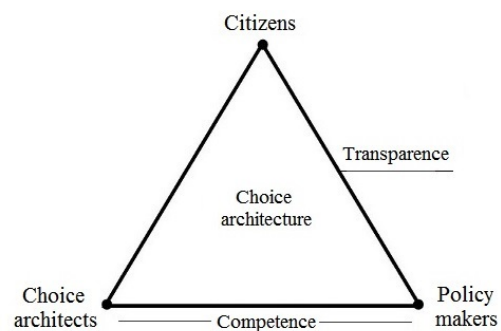


Figura 1. La geometria tra cittadini, architetti della scelta e *policy makers* (fonte?)

Nel caso del *Behavioural Insights Team* (*BIT*) questo rapporto triadico è ben evidente. Il *BIT* è un collettivo di economisti comportamentali che dal 2010 collabora con diversi dipartimenti governativi britannici, come il NHS (*National Health Service*) e il DVLA (*Driver and Vehicle Licensing Agency*). Il team opera applicando un *framework* (racchiuso nell'acronimo *MINDSPACE*⁵), in cui vengono separate e studiate le varie componenti che influenzano uno specifico problema pubblico e sociale da risolvere. In questo modo ha condotto importanti ricerche nell'ambito del pagamento delle tasse e dell'evasione fiscale (BIT, 2012a), delle donazioni caritatevoli (BIT, 2013), del sistema sanitario (BIT, 2011a), del risparmio energetico (BIT, 2011b) e dell'aumento di potere dei consumatori (BIT, 2012c).

Nel tentativo di migliorare il sistema di ricollocamento nel mercato del lavoro, la "Nudge Unit" è riuscita a creare un programma fondato su alcuni assunti compor-

tamentali, condensati in tre direttive: (a) essere sicuri che il cliente che ha perso il lavoro inizi a parlare di una sua possibile nuova assunzione subito nei giorni seguenti al licenziamento, e non due settimane dopo, come di prassi; (b) introdurre documenti in cui si richieda l'impegno del cliente a concentrarsi sin nei quindici giorni successivi nella ricerca, divisa per semplici passi, di un nuovo impiego, in collaborazione con il responsabile dell'ufficio di collocamento; (c) costruire una forma di elasticità psicologica e di benessere per coloro che, dopo otto settimane, siano ancora alla ricerca di un impiego. Gli effetti ammontano a un aumento del 15-20% nel numero di cittadini in grado di ritrovare un lavoro nell'arco di tre mesi (BIT Blog, 2012).

Un eccellente esempio di ente di governo che, collaborando con architetti delle scelte qualificati e competenti, può agire sull'architettura delle scelte di chi ha perso il proprio lavoro, e tramite un percorso trasparente e innovativo, rendere un servizio ai cittadini con risultati comprovati.

Su questa base, suggeriamo di catalogare i *nudge* attraverso l'utilizzo di tre diverse proprietà continue, non-discrete, che caratterizzano ogni singolo intervento⁶:

(x) il costo in termini di *elaborazione cognitiva* da parte del soggetto⁷;

⁵ Gli individui sono fortemente influenzati da chi trasmette l'informazione (*Messenger*); la loro risposta a incentivi (*Incentives*) economici è regolata da diversi meccanismi cognitivi; sono condizionati da ciò che fanno gli altri (*Norms*); tendono a scegliere opzioni abitudinarie o già impostate (*Default*); la loro attenzione è diretta verso elementi circostanti nuovi e più rilevanti (*Salience*); le loro azioni sono spinte da segnali che agiscono a livello inconscio (cioè automatico) presenti nell'ambiente (*Priming*) e dagli stati emotivi (*Affect*); cercano di essere coerenti con le loro promesse pubbliche e con gli atti di reciprocità (*Commitment*) e agiscono per migliorare la propria autostima e il proprio benessere (*Ego*) (Dolan et al., 2012).

⁶ Una proposta simile è presente in Ly et al. (2013), dove però vengono utilizzate quattro caratteristiche non continue e dove non si fa riferimento alla centralità del processo di "trasparenza".

⁷ Hansen e Jespersen (2013) distinguono concettualmente "scelta" e "comportamento" ponendoli

(y) il grado di motivazione e *volontà* da parte del soggetto nel compiere l'azione che si intende incentivare⁸;

(z) il livello di *trasparenza* dell'intervento.

Un *nudge* occupa un punto preciso nello spazio cartesiano dove ognuna delle tre dimensioni corrisponde alle sopracitate caratteristiche (Figura 2). In questo modo è possibile definire il conseguente grado di *accettabilità* degli interventi, che dipende in ultima istanza dalla loro *visibilità*.

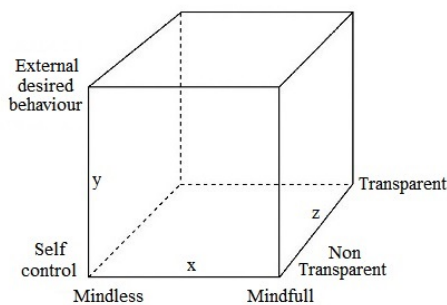


Figura 2. Il piano cartesiano a tre dimensioni del *nudge* (fonte?)

Consideriamo la famiglia delle *green defaults*, ovvero le opzioni di default utilizzate

agli estremi del continuum dell'elaborazione cognitiva. La prima coincide con il risultato di un'elaborazione riflessiva e cosciente, il secondo corrisponde agli output di processi cognitivi automatici e subpersonali.

⁸ Alcuni comportamenti, come risparmiare denaro o iniziare una dieta, possono ricevere un sostegno interno attivo, di self-control, maggiore rispetto a comportamenti che possono risultare indifferenti al soggetto, come il pagamento delle tasse, e quindi più propensi ad essere indirizzati dall'esterno. I *nudge*, dunque, possono servire per ampliare, e non limitare, la *libertà di scelta* di un individuo, aumentandone il self-control e supportando il raggiungimento di obiettivi auto-imposti.

per promuovere l'utilizzo di apparecchi volti ad un risparmio energetico. In due città della Germania, i tassi di adesione a dispositivi per le energie verdi si aggirano intorno al 90%. Questo per merito del fatto che nei rifornimenti di energia elettrica ai cittadini e nella scelta dei piani tariffari corrispondenti è pre-selezionato l'utilizzo di energia proveniente da fonti rinnovabili. Lo stesso vale per l'installazione di contatori intelligenti, che forniscono ai consumatori un feedback in tempo reale sui propri consumi (Prichert, Katsikopoulos, 2007; Kaenzig et al., 2013; Sunstein, Reisch, 2013). In questo caso è chiaro come un intervento i cui benefici sono trasparenti (z) sia in grado di attivare un comportamento desiderato dalla comunità (y), con ricadute virtuose sul singolo individuo, al quale non viene imposto alcun costo cognitivo (x), vista la semplicità con cui è possibile aderire all'opzione di default.

Semplicità che è diventata la parola d'ordine nei primi progetti di *smart disclosure*, che negli Stati Uniti e nel Regno Unito permettono ai consumatori di far fronte all'asimmetria informativa tra i servizi di cui usufruiscono e i propri consumi. Dalla sottoscrizione di mutui a quella di contratti telefonici, i cittadini conoscono poco se stessi, ignorando o sopravvalutando le proprie reali esigenze (Kamenica, Mullainathan, Thaler, 2011). Si pensi alle spese bancarie, assicurative, sanitarie ed energetiche, spesso oscure, forfettarie e fornite con tempistiche slegate dal consumo. In un enorme processo di *democratizzazione dei dati*, questi ultimi per la prima volta sono scaricabili in qualsiasi momento in un formato elettronico universale.

Questo permette di poterli caricare su motori di ricerca apposti, che trovano il servizio che piú soddisfa il consumatore in termini di rapporto qualità-prezzo-esigenze, semplificando e rendendo piú informate le scelte dei cittadini (Loewenstein, Sunstein, Goldman, 2013).

Interventi innovativi di questo tipo, semplici e non dispendiosi, raggiungono dunque risultati migliori rispetto a metodi tradizionali. Si pensi alle campagne di sensibilizzazione per il rispetto del decoro pubblico. In alcune zone di Copenhagen adesivi sull'asfalto, raffiguranti impronte di piedi verdi, conducono ai bidoni della spazzatura. Aumentare la salienza visiva di questi ultimi (processo che richiede poca elaborazione cognitiva) ha portato ad una diminuzione del 46% del fenomeno di gettare rifiuti a terra con noncuranza (Jespersen, 2012).

3.1 Varianza culturale

Conoscere le influenze che operano sul comportamento degli individui è solo un primo passo. Quello successivo è avere un riferimento, strumenti precisi, con cui individuare interventi specifici per determinati problemi (Johnson et al., 2012). Questo è possibile conoscendo ciò che è stato fatto, ovvero ciò che *ha funzionato* in qualche parte del mondo, e ciò che è *riproducibile* e ideabile a partire dalle conoscenze nell'ambito psicologico e comportamentale. E' necessario inoltre considerare la *varianza semantica*: le percezioni di un significato possono variare al modificarsi di culture diverse

o delle risposte di individui all'interno dello stesso contesto (Selinger, Whyte, 2010; 2011). Perciò va tenuta in primo piano l'interazione tra proprietà e variabili culturali, poiché esse causano differenti suscettibilità ai *nudge* sottostanti. E' piuttosto ovvio che, per esempio, le elaborazioni percettive (es: adesivi posti sull'asfalto, raffiguranti false buche o dossi, usati per rallentare la velocità) siano resistenti al variare delle culture, al contrario di comportamenti piú complessi (es: si pensi alla diversa tendenza al risparmio tra occidentali ed asiatici) (Bovens, 2010).

Distinguere la varianza culturale nelle *applicazioni* della spinta gentile da quella dei suoi *obiettivi* porterà a risultati migliori e piú cogenti, a *real nudges*, operativi e contestualizzati nel mondo reale, e non solo nei laboratori (Bovens, 2012).

3.2 Valutazione del *nudge*: un circolo virtuoso

Per avere *nudges* efficaci nelle situazioni del mondo reale e non solo in laboratorio può essere utile delineare quella che chiameremo *prassi del nudge*, individuabile in tre fasi differenti:

- (I) *definire* il problema su cui agire, e mappare il contesto di scelta degli individui, comprendendo quali meccanismi cognitivi di decisione entrino in gioco;
- (II) pensare al *design* e all'implementazione dell'intervento cognitivo-

comportamentale, tenendo conto delle caratteristiche viste sopra ed individuando i possibili impedimenti e le soluzioni che più si avvicinano ad un grado di *trasparenza* accettabile;

(III) passare alla *valutazione* di efficacia dell'intervento.

Una serie di ricerche volte a incrementare le donazioni ad associazioni caritatevoli semplifica con efficacia questo procedimento.

Gli interventi si sono basati su quattro assunti comportamentali, sperimentalmente consolidati. Le donazioni devono essere *facili*, prive di impedimenti, e *attraenti*, per risaltare la dimensione *sociale* dell'azione, che va incentivata nei momenti di scelta maggiormente opportuni⁹. Si è deciso, attraverso un nuova opzione di default, che i donatori aumentassero periodicamente i propri versamenti futuri, in maniera tale che non venissero erosi dall'inflazione. Sono stati utilizzati messaggi personalizzati per dare ampia visibilità e pubblicità ai donatori virtuosi. Introdurre un'immagine di questi ultimi ha raddoppiato i tassi di partecipazione. Si è inoltre verificato che le persone sono più propense a donare nel mese di dicembre, rispetto che a gennaio, e tendono a preferire aumenti automatici delle elargizioni future piuttosto che presenti. In questo modo si sono migliorati del 49% i risultati ottenuti con metodi tradizionali (BIT, 2013).

⁹ Questi quattro elementi valgono in qualsiasi tentativo comportamentalmente informato per incentivare un'azione. Il BIT ha ribattezzato questo *framework* **EAST**: *Easy, Attractive, Social, Timely*.

Com'è possibile valutare gli output di uno specifico *nudge*? In letteratura si tende a lasciare implicita l'analisi riguardante la fase (III). Il processo di implementazione viene presentato come un percorso lineare, al termine del quale si può avere un esito positivo o negativo.

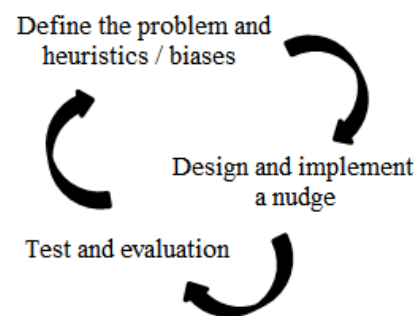


Figura 3. Il processo di implementazione virtuosa del *nudge* (fonte?)

In realtà, la descrizione di un'attenta prassi del *nudge* è meglio rappresentata da uno sviluppo circolare (Figura 3). I dati ottenuti devono essere utilizzati per arricchire e migliorare le fasi (I) e (II). Si parte da un'ipotesi d'intervento (*ex-ante*), fatta di assunti corroborati da studi in laboratorio e sul campo. In seguito all'implementazione di uno specifico intervento, queste ipotesi possono essere messe in dubbio o rinforzate dai risultati ottenuti dalla valutazione (*ex-post*) di efficacia. Questa avviene attraverso il metodo sperimentale, che la nascente *evidence-based policy* vorrebbe estendere a ogni processo di *policy making*. Una pratica consapevole del *nudge* deve dunque affrontare la questione di come unire la propria metodologia con la costante ricerca dell'evidenza nei risultati ottenuti.

4. *Policy making* di provata efficacia

Negli ultimi anni, un numero sempre maggiore di discipline accademiche ha incoraggiato l'applicazione del metodo sperimentale al *policy making*. In questo modo è possibile capire quali progetti funzionino e diminuire il grado di incertezza che caratterizza ogni intervento calato nella complessità del mondo sociale. L'*evidence-based policy* basa la pratica del *policy making* sulla provata efficacia.¹⁰ (Sutcliffe, Court, 2005; Ludwig, Kling, Mullainathan, 2011; Liebman, 2013). In questo modo si può sottrarre il processo di progettazione e implementazione di politiche economiche, pubbliche e sociali dal dominio delle intuizioni, dei dogmi e dei pregiudizi, che tipicamente caratterizzano il dibattito ideologico. Questo rinnovamento è dovuto sia alla trasformazione dell'economia, molto più aperta ad approcci sperimentali (Guala, 2005; Charness, Gneezy, Kuhn, 2013), sia al successo ottenuto dalle scienze mediche (Howick, 2011).

Proprio da queste ultime viene preso in prestito il termine *evidence-based*, per indicare quegli interventi di cura e prevenzione basati su risultati rigorosi di test clinici, cioè sui *randomised controlled trials* (RCTs). La parola *trials* denota il carattere sperimentale dello

studio e di come lo sperimentatore giochi un ruolo fondamentale nella decisione dei parametri del set d'intervento. L'aggettivo *controlled* fa riferimento alla pratica di prendere in considerazione il confronto dei risultati dati da due (o più) gruppi di individui (o popolazioni, aree geografiche, etc). Nel gruppo di sperimentazione, viene applicato il trattamento che si intende studiare, mentre un gruppo di controllo ne riceve uno diverso o nessuno. E' necessario che i due gruppi siano il più omogenei possibili e che differiscano, per l'appunto, solo per la variabile che si intende manipolare. L'aggettivo *randomised*, invece, determina che l'assegnazione degli individui ai diversi gruppi debba avvenire attraverso un metodo casuale, in maniera tale da permettere una distribuzione omogenea delle variabili. La randomizzazione è dunque fondamentale almeno per due motivi: evita che interessi personali e giudizi volti a sfavore un intervento piuttosto che un altro possano invalidare l'esperimento, e fornisce una solida base per i metodi standard dell'analisi statistica.

Su questa metodologia si fonda la *valutazione sperimentale* della maggior parte delle ricerche fino ad oggi condotte nell'ambito degli approcci comportamentali, tra cui quelle avviate e implementate dal *BIT*¹¹. Dall'utilizzo di messaggi sociali per la riscos-

¹⁰ Alcuni autori preferiscono l'espressione *evidence-informed policy* (Chalmers, 2005; Sharples, 2013). Negli Stati Uniti e nel Regno Unito sono presenti organizzazioni promotrici di questo nuovo approccio (si vedano, tra le tante, la *Coalition for Evidence-Based Policy* e l'*Alliance for Useful Evidence*).

¹¹ Il metodo sperimentale del *BIT* è riassumibile in tre fasi: a) *Test*, creare un set di sperimentazione valido; b) *Learn*, misurare i risultati dell'intervento; c) *Adapt*, utilizzare quest'ultimi per arricchire gli assunti iniziali con cui si è iniziata la valutazione (Behavioural Insights Team, 2012b).

sione delle tasse al tentativo di creare nuovi programmi per il reinserimento del mondo del lavoro, tutto si basa sull'idea di isolare due gruppi, applicando a uno l'intervento e all'altro no. I risultati raccolti, se differenti, vanno imputati all'unica variabile manipolata, all'applicazione di un *nudge* di un qualche tipo. Questa valutazione di efficacia e comprensione dei fattori casuali in campo può avvenire su scale di grandezza differenti, a seconda degli strumenti a disposizione e della complessità del comportamento che si vuole influenzare.

Entrambi gli estremi sono possibili. Si possono condurre *studi controllati randomizzati* su un ristretto numero di partecipanti, i quali hanno per esempio mostrato la ripetuta tendenza a scegliere nei menù dei ristoranti i cibi posti alle estremità della lista, con la conseguente possibilità di implementare una spinta gentile che aiuti il consumo di piatti maggiormente salutari (Dayan, Bar-Hillel, 2012). Oppure si può agire con trials di enorme portata. In una recente ricerca, tramite il social network Facebook, negli Stati Uniti è stato possibile isolare durante una tornata elettorale tre gruppi di utenti, per un ammontare di 61 milioni di contatti. Mentre alla maggior parte di quest'ultimi veniva ricordato il fatto che si tenessero le elezioni al Congresso, a 600.000 utenti veniva fatta presente la lista di amici che erano andati già a votare (ad un terzo gruppo di controllo non veniva applicato alcun trattamento). Con il secondo, basato sull'influenza del fattore sociale, si sono ottenuti gradi di adesione al voto quattro volte maggiori (Bond et al., 2012).

4.1 I limiti della randomizzazione

Il ricorso alla randomizzazione come *gold standard* nelle scienze sociali non è esente da problemi epistemologici e pratici (Colyvan, La Caze, 2009). Una possibile difficoltà riguarda se davvero i *RCTs* permettano di valutare l'impatto di un programma soltanto attraverso il confronto tra due gruppi di individui, focalizzandosi solo sugli effetti e non sulla reale comprensione delle cause (Harrison, 2011). Nonostante l'uso del metodo casuale, eliminare i più comuni *bias*, come quelli dell'assegnazione e della selezione, presta il fianco al "problema della generalizzabilità". Ovvero dimostrare come si possa passare dal "ha funzionato qui" al "funzionerà anche lì". Se gli *RCTs* si fanno carico di offrire questa capacità alle istituzioni politiche, non rendono conto della loro mancanza di *validità esterna*, basandosi solo su una robusta *validità interna* (Cartwright, Hardie, 2012). Le valutazioni randomizzate non ci forniscono solidi modelli causali, poiché sono inficiati dal fatto che mancano della potenzialità di replicare i propri risultati in altri contesti. Esse garantiscono solo i risultati di *quell* esperimento, non di tutti gli altri che seguono il medesimo schema (Worrall, 2010; Rousch, 2008).

L'obiezione è giustificata, ma non insuperabile sul piano pratico. Compito futuro sarà puntare sulla *replicabilità* e *riproducibilità* degli esperimenti, riuscendo di volta in volta a isolare le *cause* che influenzano certi comportamenti e che permettono di incentivarne altri. Da un punto di vista *pragmatico*, i risultati non andranno presi sempre come certi e

generalizzabili in ogni contesto, ma rivisti dal normale percorso di ricerca dell'evidenza, sostenuti dal *framework* teorico delle scienze comportamentali, sempre più robusto e strutturato in teorie di più alto livello di generalità.

Sul piano etico, avere a che fare con individui pone il problema di come possa ritenersi moralmente accettabile l'utilizzo di criteri randomizzati per attribuire trattamenti ad alcuni ed altri no. Studi nel campo dell'economia nei paesi in via di sviluppo hanno mostrato, tramite la randomizzazione, come semplici iniziative che tengano conto della *razionalità limitata* degli individui possano avere grande successo nell'aumentare i tassi di vaccinazione, di adesione alla scuola dell'obbligo o all'utilizzo di fertilizzanti (Banerjee, Duflo, 2012; Datta, Mullainathan, 2012) Appare però arbitrario e non egualitario, per esempio, fornire forme di prevenzione solo per certi gruppi di persone e ad altri no. In realtà, oltre alla limitatezza dei fondi, prima della valutazione, bisogna tener conto che non si conosce davvero quale sia l'intervento migliore tra i due (o più) testati. Perciò sembra ancor più eticamente scorretto continuare a poggiare il *policy making* sull'ignoranza di cosa funzioni meglio¹².

Politiche pubbliche basate sull'evidenza non possono dunque prescindere dalla riflessione su che cosa sia in ultima istanza l'*evidenza* stessa. Come possono alcuni fatti

¹² Nella lotta alla povertà, le organizzazioni leader, che uniscono ricerche comportamentali con la raccolta di *evidenza*, sono *ideas4*, *Innovation for Poverty Action* e il *Poverty Action Lab*.

candidarsi come essi stessi evidenti? Come facciamo a decidere quanta evidenza abbiamo e come facciamo a valutarla? Come si possono combinare evidenze diverse, provenienti magari anche da metodi diversi? Sono interrogativi legittimi e importanti che richiedono una soluzione. Ma occorre anche tenere conto che i *RCTs* sono una delle possibili fonti di evidenza, e non esauriscono il dominio di quest'ultima. Un ruolo importante va assegnato anche agli altri stadi delle *gerarchie dell'evidenza*, come le meta-analisi e le indagini qualitative¹³.

5. Conclusione

Gli approcci *behaviorally-informed* sono solo uno dei possibili approcci all'*evidence-based public policy*, ma hanno al loro fianco un grandissimo apporto di originalità. Architetture delle scelte intelligenti, a misura di cittadino, sono le alternative a contesti dominati da regole ed imposizioni esplicite, con l'appetibile conseguenza di poter finalmente fare *in parte* a meno della regolamentazione come strumento primario di orientamento del comportamento (Barr, Mullainatha, Shafir, 2009). Sul piano *empirico*, questa nascente prospettiva ha già mostrato i suoi primi successi. Scopo delle ricerche future è sviluppare un quadro teorico e operativo, metodologicamente

¹³ Recentemente l'idea di gerarchia è stata messa in questione, in favore di un'evidenza "multidimensionale" (Nutley, Powell, Davies 2013), o distribuita in una struttura a rete (Bluhm, 2005).

consapevole, che possa svolgere il ruolo di guida e di cartina tornasole per i *policy makers*. Affinchè chi progetta ambienti di scelta, più semplici, sostenibili e vantaggiosi per i cittadini e la società possa farlo in modo consapevole e responsabile.

Bibliografia

- Allcott, H. (2011), "Social norms and energy conservation", *Journal of Public Economics*, 95, 9-10, 1082-1095
- Banerjee A., Duflo E., (2012), *Poor Economics. A Radical Rethinking of the Way to Fight Global Poverty*, Public Affairs, New York
- Barr M., Mullainathan S., Shafir E. (2009), "The Case Of Behaviorally Informed Regulation", in Cisternino J. (ed.), *New Perspectives in Regulation*, Cambridge, MA
- Behavioural Insights Team, (2011a), "Applying behavioural insights to health", Cabinet Office
- Behavioural Insights Team, (2011b), "Better Choices: Better Deals", Cabinet Office
- Behavioural Insights Team (2011c), "Behaviour change and energy use", Cabinet Office
- Behavioural Insights Team (2012a), "Applying behavioural insights to reduce fraud, error and debt", Cabinet Office
- Behavioural Insights Team (2012b), "Test, Learn, Adapt", Cabinet Office
- Behavioural Insights Team (2012c, December 14), "New BIT trials results: helping people back into work", blogs.cabinetoffice.gov.uk/behavioural-insights-team/2012/12/14/new-bit-trials-results-helping-people-back-into-work
- Behavioural Insights Team (2013), "Applying behavioural insights team to charitable giving", Cabinet Office
- Benartzi S., Thaler R., (2013), "Behavioral Economics and the Retirement Savings Crisis", *Science*, 339, 6124, 1152-1153
- Bluhm R., (2005), "From Hierarchy to Network: A Richer View of Evidence from Evidence-Based Medicine", *Perspectives in Biology and Medicine*, 48, 4, 535-547
- Bond R. M., et al., "A 61 Million Person Experiment In Social Influence And Political Mobilisation", *Nature*, Research Letter, 489, 295-298
- Bovens L., (2010), "Nudges and Cultural Variance: a Note on Selinger and Whyte", *Know Techn Pol*, 23:483-486
- Bovens L., (2012), "Real Nudge", *European Journal of Risk Regulation*, 1
- Brooks D., (2013), "The Nudge Debate", *The New York Times*, <http://nyti.ms/14saET3>
- Cartwright N., Hardie J., (2012), *Evidence-based Policy: A practical guide to doing it better*, Oxford University Press
- Chalmers I. (2005), "If Evidence-informed Policy Works In Practice, Does It Matter If It Doesn't Work In Theory?", *Evidence & Policy*, 1, 2, 227-442
- Charness G., Gneezy U., Kuhn M. A., (2013), "Experimental Methods: Extra-Laboratory Experiments-Extending the Reach of Experimental Economics", *Journal of Economic Surveys*, 27, 1, 1-50

- nal of Economic Behavior and Organization*, 91, 93-100
- Chetty R., et al. (2012), "Active Vs. Passive Decisions And Crowd-out In Retirement Savings Accounts", NBER Working Paper No. 18565
- Chetty R., et al. (2013), "Subsidies vs. Nudges: Which Policies Increase Saving the Most?", *Center for Retirement Research*, 13-3
- La Caze A., Colyvan M., (2009), "Evidence-based Policy- Promises And Challenges", unpublished manuscript
- Datta S., Mullainathan S., (2012), "Behavioral Design. A new approach to development policy", CGD Policy Paper 016
- Dayan E., Bar-Hillel M., (2012), "Nudge To Nobesity II: Menu Positions Influence Food Orders", *Judgment and Decision Making*, 6, 4, 333-342
- Dolan P., et al., (2012), "Influencing Behaviour: The Mindspace Way", *Journal of Economic Psychology*, 33, 264-277
- Fryer R., et al., (2012), "Enhancing The Efficacy Of Teacher Incentives Through Loss Aversion. A Field Experiment", NBER Working Paper No. 18237
- Goldstein D., et al., (2008), "Nudge Your Costumers Toward Better Choices", *Harvard Business Review*
- Guala F., (2005), *The Methodology of Experimental Economics*, Cambridge University Press
- Hansen P. G., (2012), "Should We Be 'Nudging' for Cadaveric Organ Donations?", *American Journal of Bioethics*, 12, 2, 46-48
- Hansen P. G., Jespersen A. M. (2013), "Nudge and manipulation of choice. A framework for the responsible use of the nudge approach to behaviour change in public policy", *European Journal of Risk Regulation*, 1, 3-28
- Harrison G., (2011), "Randomisation And Its Discontents", *Journal of African Economies*, 20, 4, 626-652
- Hausman D., Welch B., (2010), "Debate: To Nudge or Not to Nudge", *The Journal of Political Philosophy*, 18, 1, 123-136
- Howick J., (2011), *The Philosophy of Evidence-based Medicine*, Wiley-Blackwell, BMJ Books
- Jespersen, S.M. (2012, February 16). "Green nudge: Nudging into the litter bin" <http://www.inudgeyou.com/green-nudge-nudging-litter-into-the-bin>
- Johnson E. J., Goldstein D., (2003), "Do Defaults Save Lives?", *Science*, 302, 5649, 1338-1339
- Johnson E. J., et al. (2012), "Beyond nudges: Tools of choice architecture", *Marketing Letters*, 23, 487-504
- Kaenzig J., et al., (2013), "Whatever the Customer Wants, the Customer Gets? Exploring the Gap between Consumer Preferences and Default Electricity Products in Germany", *Energy Policy*, 53, 311-322
- Kalbekken S., Saelen H., (2013), "Nudging Hotel Guests To Reduce Food Waste As A Win-win Environmental Measure", *Economics Letters*, 119, 325-327
- Kamenica E., (2012), "Behavioral Economics And Psychology Of Incentives", *Annual Review of Economics*, 4, 13.1-13.26

- Keller P. A., (2011), “Enhanced Active Choice: A New Method To Motivate Behavior Change”, *Journal of Consumer Psychology*, 21, 4, 376-383
- Institute for Government (2010), *MINDSPACE: Influencing behavior through public policy*
- Liebman J., (2013), “Building on Recent Advances in Evidence-Based Policymaking”, The Hamilton Project
- Loewenstein G., Sunstein C. R., Golman R., (2013), “Disclosure: Psychology Changes Everything”, *Annual Review of Economics*, in press
- Ludwig J., Kling J., Mullainathan S., (2011), “Mechanism Experiments and Policy Evaluation”, NBER Working Paper No. 17062
- Ly K., et al., (2013), “A Practitioner’s Guide To Nudging”, Research Report Series: Behavioural Economics in Action, Rotman School of Management, University of Toronto
- Montuschi E., (2009), “Questions Of Evidence In Evidence-based Policy”, *Axiomathes*, 19, 425-439
- Nutley S., Powell A., Davies H., (2013), “What Counts as a Good Evidence”, Report for the Alliance for Useful Evidence
- Pichert D., Katsikopoulos K., (2007), “Green Defaults. Information Presentation And Pro Environmental Behaviour”, *Journal of Environmental Psychology*, 28, 63-73
- Roush S., (2008), “Randomised Controlled Trials And The Flow Of Information. Comment On Cartwright”, *Philosophical Studies*, 143, 137-145
- Rozin P., et al. (2011), “Nudge To Nobesity I. Minor Changes In Accessibility Decrease Food Intake”, *Judgment and Decision Making*, 1, 4, 323-332
- Schultz P., Khazian A., Zaleski A., (2008), “Using Normative Social Influence To Promote Conservation Among Hotel Guests”, *Social Influence*, 3, 1, 4-23
- Selinger E., Whyte K. P., (2010), “Competence and Trust in Choice Architecture”, *Know Tech Pol*, 23, 461-482
- Selinger E., Whyte K. P., (2011), “Is there a right way to nudge?”, *Sociology Compass*, 5/10, 923- 935
- Selinger E., Whyte K. P. (2012), “Nudging Cannot Solve Complex Policy Problems”, *European Journal of Risk Regulation*, 1
- Shafir E., (2012), *The Behavioural Foundations of Public Policy*, Princeton University Press
- Sharples J. (2013), “Evidence for the Frontline”, Report for the Alliance for Useful Evidence
- Sunstein C. R. (2011), “Empirically Informed Regulation”, *University of Chicago Law Review*, 78, 4, Harvard Public Law Working Paper, 13-03
- Sunstein C. R. (2012). “Impersonal Default Rules vs. Active Choices vs. Personalized Default Rules: A Triptych”, *Regulatory Policy Program Working Paper RPP-2012-17*.
- Sunstein C., (2013), *Simpler. The Future of Government*, Simon & Schuster
- Sunstein C. R., Reisch L. (2013a), “Automatically green. Behavioral Economics and Environmental Protection”, Preliminary draft available

- at <http://ssrn.com/abstract=2245657> SSRN:
- Sunstein C. R., Reisch L. (2013b), “Green by default”, *Kyklos*, 66, 3, 398, 402
- Sutcliffe S., Court J., (2005), “Evidence-based Policy Making: What Is It? How Does It Work? What Relevance For Developing Countries?”, *Overseas Development Institute*
- Thaler R., Sunstein C. R. (2008), *Nudge. Improving Decisions about Health, Wealth and Happiness*, New Haven, Yale University Press
- Volpp K. G. et al., (2008), “Financial incentive-based approaches for weight loss: a randomized trial”, *The Journal of the American Medical Association*, 300(22):2631-7
- Volpp K. G. et al., (2009), “A Randomized, Controlled Trial of Financial Incentives for Smoking Cessation”, *The New England Journal of Medicine*, 360:699-709
- White M. D., (2013), *The Manipulation of Choice: Ethics and Libertarian Paternalism*, Palgrave Mac Millan
- Worrall J., (2010), “Evidence: Philosophy Of Science Meets Medicine”, *Journal of Evaluation in Clinical Practice*, 16, 356-362

Theories of Team Reasoning

.....
Alessandra Smerilli

Università Pontificia Salesiana
(Italia)

1. Introduction

The idea of we-thinking, or we-reasoning, is increasingly drawing the attention of more and more economists. In its general formulation, it has been proposed by David Hodgson (1967), Donald Regan (1980), Margaret Gilbert (1989), Susan Hurley (1989), Raimo Tuomela (1995), and Martin Hollis (1998). Within this body of literature, Robert Sugden (1993, 2000, 2003) and Michael Bacharach (1995, 1997, 1999, 2006) have developed analytical frameworks from an economic point of view. The main claim of scholars that analyze we-thinking is that it can be endorsed by people when they face a decision problem. In fact, experimental evidence shows that, especially in some kind of games, such as coordination games, people do endorse we-thinking¹. However, the way in which we-thinking arises and how it brings people to behave in a particular way in games is a matter that requires further investigation.

¹ See Tan and Zizzo (2008) for a review of experiments.

The two main contributors are Bacharach and Sugden, and they approach the topic in two different ways. Sugden's aim is to show that we-reasoning is a consistent and logical way of thinking, but he does not face the problem of how we-reasoning can arise. He gives some intuitions only about a psychological background based on Smith's analysis of correspondence of sentiments.

Bacharach's theory is based on frames and his never reached aim (because of his death) was to explain we-thinking in terms of Variable Frame Theory (Bacharach 1993). But, as we shall see later, some of his intuitions conflict with the logical analysis he proposes. His logical analysis was developed by Zizzo and Tan (2003, 2008): They introduce a 'game harmony' measure, as a proxy of the extent of cooperation or conflict in games. Game harmony could represent a step forward towards the endogenization of the probability that we-thinking could arise in a decision problem.

In the present paper I offer a review of the most known theories of we-thinking and I focus in particular in Bacharach's theory.

2. Theories of Team Reasoning

All the existing theories of team reasoning share the idea that people can group-identify in front of a decision problem and that they try to reach the best outcome for the whole group, doing their part in the best combination of actions.

This kind of reasoning is called team or we-reasoning and it is the effect of group identification.

Theories of we-reasoning are basically divided into two kinds of approaches: some scholars consider group formation as rational, others do not.

For Bacharach, for example, team-reasoning is a result of a psychological mechanism – the group identification; for Gilbert (1989) and Tuomela (1995) group formation is a result of a mutual commitment; Regan (1980), in his “Utilitarianism and co-operation” proposes a normative theory for moral and rational agents in which the rule to follow is: “What each agent ought to do is to cooperate, with whoever else is cooperating, in the production of the best consequences possible given the behaviour of non cooperators” (p.124).

Susan Hurley offers a theory of ‘rationality’ of we-thinking. Defining the unit of agency as “the unit the causal consequences of the activity of which are in question” (1898, p.140), Hurley identifies as units the subsystem (‘each’) or the system (‘we’) and claims that those units have not to be taken as fixed.

In fact, in front of a decision problem an agent firstly must ask herself: which is the objective in this situation? Subsequently She can choose the unit of agency, which is the most appropriate for the objective: “An adequate theory should help us to understand what the appropriate unit of agency is in various circumstances” (p.146). The consequence of this reasoning is that it is rational to allow different units of agency, hence we-

thinking is also rational. A question, however, remains open: “If units of agency are not exogenously fixed, how are units formed and selected? Is centralized information or control required, or can units emerge as needed from local interactions? At what points are unit formation and selection rationally assessable?” (2003, p.165). The last question is very important, because the fact that unit formation is rational has still to be demonstrated. In fact, as Hollis and Sugden point out, in answering the question whether the formation of a unit is a requirement of rationality: the question whether the formation of a unit is a requirement of rationality: “If

we are to stay at all close to account of rationality that derives from Hobbes, Hume, Bentham, Pareto and Savage, we must answer ‘No’ (Hollis and Sugden 1993, p. 13)”. In fact, they explain that according to their account of rational choice theory, a choice is rational in relation to the desires or preferences of the agent who is making the choice: “a choice can be rational only for a particular agent” (ib.). It follows that a theory of rationality cannot give an account of the formation of the unit of agency. Differently from Hurley, who claims that there must be agent neutral goals to be pursued, Elisabeth Anderson (2001) states that the determination of personal identity, which can be plural or individualistic, precedes the choice of the kind of reasoning to be adopted. She makes use of team reasoning in order to give an account of the ‘rationality of committed action’: “regarding themselves as members of a single collective agency, the parties are

committed to acting only on reasons that are universalizable to their membership” (p.29). She then states the ‘Priority of Identity Rational Principle’: “what principle of choice it is rational to act on depends on a prior determination of personal identity, of who one is” (p.30). Following the previous principle, Anderson shows that either acting on maximization of expected utility or on team reasoning is a rational act, depending on regarding oneself as an isolated individual or a member of a team. In Anderson’s account, then, the determination of personal identity comes before the decision of what principle of choice is in play.

In the following section Bacharach’s theory will be analysed.

3. Bacharach’s theory of we-thinking

The answers to fundamental questions about coordination and cooperation ... lie in the agent’s conception not of the objects of choice, nor of the consequences, but of herself and of the agents with whom she is interacting (Bacharach, M., N. Gold and R. Sugden 2006, p. 70).

This sentence is the starting point of Bacharach’s analysis of we-reasoning².

² It is interesting to see how and when Bacharach developed the idea of we-thinking. He started by building the Variable Frame theory (Bacharach 1993), in parallel he was developing a theory of cooperation. In 1995 he introduced the category of ‘fellow member reasoner’: “Someone who is a

member of a natural type T and chooses a certain strategy

if she is sufficiently sure that her interactants are also member of T” (1995, p.1). In this context he tries to link T-membership to variable frame theory and, at for the first time he introduces the ‘we’ category: “The present paper has made type T membership an issue which type T members think about, and nuanced their capacity to recognize it. An alternative development would make T membership a variable element in players’ frames in the sense of variable frame theory: that is, a player might or might not think about the game in terms of whether she and her coplayers belong to T. In the case in which T is the player set, we may put this by saying that a player may or may not think in ‘we’ terms about how to play the game. The more inclined a player is to ‘we’ thinking, and the more inclined she takes coplayers to be, the more will fellow-member reasoning be favoured” (p.17). In 1997 Bacharach formally introduces we-thinking, but, at the same time, he claims that we-reasoning is different from team-thinking: “the essence of team thinking is that each player participates in the group-best profile in conditions of common knowledge that they form a ‘team’ - that is, that they all act in this way” (p.13). The theory of ‘team-reasoning’ he has in mind was proposed by Robert Sugden in 1993. This was a seminal paper in Sugden’s research project on we-thinking. In fact later versions of his theory (see Sugden 2000, 2003) do not assume the participation of all subject to participate and the common knowledge hypothesis. The first published paper in which Bacharach formalizes his theory is the 1999’s article about ‘interactive team reasoning’. In it Bacharach introduces some elements that we can find in the book, such as, group identification, team reasoning as the effect of group identification, unreliable team interaction, that in the book becomes circumspect team reasoning, etc. Between the ‘99 article and the book we may find some lecture notes, in which the concepts of agency and ‘superagency’ begin to ap-

We-reasoning is seen as a powerful ‘mechanism’ (in Bacharach’s words) for solving puzzles about cooperation and coordination in game theory (i.e. games like Hi-Lo and Prisoner Dilemma). More in general, in his work Bacharach tries to demonstrate, by showing some evidence³, that we-reasoning is a valid mode of reasoning and people do endorse it.

His theory is based on frames: if we-frame comes to mind, the subject will group-identify and then she will start to we-reason. A frame can be defined as a set of concepts that an agent uses when she is thinking about a decision problem. It cannot be chosen, and how it comes to mind is a psychological process: “Her frame stands to her thoughts as a set of axes does to a graph; it circumscribes the thoughts that are logically possible for her (not ever but at the time). In a decision problem, everything is up for framing. . . also up for framing are her co-players, and herself”. (ib. p. 69).

In Bacharach’s framework, then, a person may start to we-reason only if she has ‘we’ concepts in her frame: in other words, a person firstly recognizes the we-perspective, and then endorses it.

pear. The book represents an (incomplete, because of his death) attempt to build a complete theory of we-thinking.

³ Bacharach claims that there are five kinds of evidence: logical, introspective, evolutionary, transcendental and experimental (see Bacharach 2006, pp.145-146).

The ‘framing’ perspective is not the only way in which the theory of we-thinking has been proposed, however. Robert Sugden, for example, has developed a different framework for looking at the problem. In his framework the central concept is the ‘common reason to believe’⁴: people who group identify are not committed to reason as a team unless there is a common reason to believe that other agents are doing the same. The psychological side of group identification in Sugden’s theory might be found in his analysis of Smith’s ‘correspondence of sentiments’ (Sugden, R. 2005): ‘fellow-feeling’ could be seen as the source of group identification. In sum, in Bacharach’s framework if people group identify they automatically start to reason like a team, whereas in Sugden’s theory people may group-identify, but team reasoning does not follow automatically.

Bacharach’s aim is to explain situations in which some people may ‘we’-reason and some other may not. In order to model the previous intuition, he assumes that a frame ‘we’ can come to mind with a probability

⁴ Common reason to believe is defined as follows: “there is common reason to believe to a proposition p in a set of individual T if: (i) for all individuals i in T, i has reason to believe p (i.e. p can be inferred from propositions that she accepts as true n.d.a.); (ii) for all individuals I and j in T, I has reason to believe that j has reason to believe p; (iii) for all individuals I, j, and k in T, I has reason to believe, that j has reason to believe that k has reason to believe p; and so on” (Gold and Sugden 2008, p. 302).

‘ ω ’, which represents the probability for a subject to group-identify. The probability ω is common knowledge⁵, in fact: “in coming to frame a situation as a problem ‘for us’, an individual also gains some sense of how likely it is that another individual would frame it in the same way” (Bacharach, Gold and Sugden 2006 p. 163). A context in which some people may group-identify and some may not is seen by Bacharach as an unreliable coordination context, and team reasoning in this context is called *circumspect team reasoning*.

In Bacharach’s theory group identification is a framing phenomenon that determines choices by “changing the logic by which people reason about what to do” (ib). If, by reasoning in the individual standard mode (I-reasoning), an agent looks at a decision problem by thinking what it would be the best for her to do, when there is group identification, the agent will think: “What would the best be for us to do?”. Basically then, “Somebody ‘team reasons’ if she works out the best feasible combination of actions for all the members of her team, then does her part in it” (Bacharach 2006 p.121).

Sugden describes team reasoning in a similar way: “The idea is that, in relation to a specific decision problem, an individual

may conceive of herself as a member of a group or team, and conceive of the decision problem, not as a problem for her but as a problem for the team. In other words, the individual frames the problem, not as ‘What should I do?’, but as ‘What should we do?’” (2000, pp. 182-183).

Trying to go deeper into the matter, Bacharach aims to explain how a frame, in particular a we-frame, may come to mind. In his earlier works (Bacharach, M. 1997, 1999) he proposes that the possibility of team reasoning is related to having ‘scope for co-operation’ and to the ‘harmony of interests’.

In his last book he suggests the (strong) Interdependence Hypothesis, that roughly states: “perceived interdependence prompts group identification”.

The perception of interdependence between two agents in a game is given by three factors:

- ✓ common interest (the agents have common interest in s^* over s , if both prefer s^* to s , where s^* , s are possible state of affairs, or, in a game, possible outcomes)
- ✓ co-power (nobody can reach s^* alone, but both can together)
- ✓ standard solution (basically Nash equilibrium that may realise s).

Bacharach says that if the two agents have common interest in s^* over s and co-power over it, and the standard solution of the game contains outcomes in s (this is the case of PD), then people are stimulated to group-identify. In other words, if s is a possible outcome of the individual rationality, and s^* is strictly Pareto-preferred by both the agents, and they have common power to

⁵ In a previous work (1999), Bacharach has developed a more formalized model, in which each agent can participate or lapse in a team and everyone, before choosing, receive a signal knowing the joint probability distribution of this signal and agent’s state (i.e. an agent’s signal includes her participation state).

bring about s^* , then group identification is stimulated. Or, if the outcome

that can be reached by an individual way of reasoning is Pareto-dominated by another outcome achievable only by thinking as a group, there is space for group identification.

The main Bacharach's purpose is to explain cooperation, seen as a successful group activity (see 2006 p. 69), and the core mechanism for doing that comprehends 'framing', 'common purpose', and 'cooperation': "(i) we frame ourselves as members of groups; (ii) . . . perceived agreement of individual goals among a set of individuals favours framing as members of a group with this common goal; (iii) the group framing tends to issue in efficient cooperation for the group goal" (p.90). People then cooperate because they group identify.

4. We-thinking and Variable Frame Theory

Bacharach's (never reached) aim was to explain we-reasoning in terms of Variable Frame Theory (VFT), which he had developed b⁶. Concisely, in VFT a player can intentionally choose an object, or an action, if she has a way of thinking about that object or that action, i.e. he has a frame. Frames can be more or less salient or available, depending on a probability measure on them. A decision

rule in VFT is "a mapping from frames to options induced by those frames" (Bacharach, M. 2001a), and an equilibrium for symmetrical games is defined as follows: "the pair (δ, δ) is a variable frame equilibrium if, for each Frame F , the option $\delta(F)$ is subjectively best from the perspective of F against δ as perceived in F " (ib.). In other words the decision rule $\delta(F)$ has to be the best reply against δ .

The intersection between VFT and we-thinking would have been called by Bacharach 'Variable agency theory' (Bacharach 2006, p.59). However, he could not complete the description of 'we'-reasoning in terms of VFT. In fact, there are at least two problems to solve, in order to complete Bacharach's theory: one is related to the way he conceives the 'we' frame, the other is the endogenization of ω .

In Bacharach's circumspect team reasoning, as I have said before, if people group-identify, then the we-frame comes to their mind and they start to we-reason. It seems as though in Bacharach's framing theory there are two aspects that are deeply linked: in framing a situation, the first step is to recognize a frame, that is coming to see it; the second step is endorsing that frame, i.e. reasoning as the frame allows you to do. In Bacharach's theory group identification means not only endorsing a particular way of reasoning, but also coming to see it. The 'compression' between the two aspects of framing is due to the VFT. In it in fact, changing frame does not mean to change the way of reasoning and the decision problem

⁶ See Bacharach 1993, 2001.

for a subject is fully determined by the interplay of his frame and the objective world.

Because of this ‘compression’, Bacharach in his theory of we-thinking cannot allow people to use more than one frame at a time. In a certain sense, as it has been noticed by Gold and Sugden (Bacharach, Gold and Sugden 2006), in we-frame people become committed to we-reason: “In the theory of team reasoning, an individual who reasons in the ‘we’ frame is aware of the ‘I’ frame too (as one of that other players might use) but acknowledge only ‘we’ reasons. It seems that group identification involves something more than framing in the sense of variable frame theory: the group-identifier does not merely become aware of group concepts, she also becomes committed to the priority of group concepts over individual ones” (p.199).

The fact that people cannot use more than one frame at a time and they cannot reason about frames whilst thinking of which frame to endorse, prevents Bacharach to develop analytically one of his intuitions. In fact, taking the most famous game in terms of cooperation, the PD game, as an example, Bacharach says: “In a Prisoner’s Dilemma, players might see only, or most powerfully, the feature of common interest and reciprocal dependence which lie in the payoffs on the main diagonal” (p.86). If this happens, players do cooperate. But, it might be the case that “they might see the problem in other ways. For example, someone might be struck by the thought that her co-player is in a position to double-cross her by playing D in the expectation that she will play C. This

perceived feature might inhibit group identification” (ib).

Here Bacharach seems to have in mind some psychological process which inhibits group identity which is not quite represented by his own concept of interdependence - the idea of ‘double-crossing’. The reason this idea does not fit his framework is that double-crossing is the incentive to act on individual reasoning when one believe the other is acting on team reasoning. This requires that the player uses both frames at the same time, while thinking about which one to use. Or, in other words, a player, in order to recognize the ‘double-cross’ threat, should be allowed to imagine himself in a we-frame, and then deliberating to cooperate, but at the same time he should use the I-frame by thinking that the other player would take advantage of her. In the first player’s conjecture, the other player too should use the we-frame in order to think that the first player could choose to cooperate, and, at the same time, she should use I-frame in order to think how ‘double cross’ the first player.

In the theory of we-thinking the way in which a person reasons (I-mode or we-mode) is a consequence of the perceived frame. So, if a person is in we-frame she cannot reason in the standard theoretical mode, and then she cannot ‘see’ the double-crossing threat. She may switch from I-mode of reasoning to we-reasoning (if we-frame comes to mind), or not. Bacharach, then, does not seem to take into account the possibility that once we are in the we-frame, we may switch to I-mode of reasoning, or

better, he allows the possibility of switching frame, but does not allow a person to be able to visualize switching frames. And this is why he cannot represent his ‘double-crossing’ intuition. It seems that when the “we” frame is perceived, it is also perceived as the correct frame or dominant frame, so that once a person sees the world this way she cannot visualize going back to seeing it the other (compare illusions, myths, lies - ‘the scales fell from my eyes’).

In effect, in one of his unpublished papers (Bacharach, M. 1997), Bacharach allowed for the possibility of the existence of three frames: the I frame, the We frame and the ‘S’ (superordinate) frame. We and I are called simple frames: “players in them begin their reasoning with the two basic conceptualization of the situation, as ‘what shall we do?’ problem and ‘what shall I do?’ problem respectively”(p. 5). A S frame is active when someone manages “during deliberation to see the problem from both the we and the I/she perspectives” (p.14). Although Bacharach allows for the existence of S, based on psychological attainments, he states that we and I perspectives cannot be held simultaneously: “Although we can switch self-identities rather easily, we appear to be unable to inhabit more than one at a time”(p.15). This seems to be a contradiction, and in fact this assumption leads Bacharach to introduce a solution concept that imposes, in his words “a stringent requirement”(p.21). He assumes that I thoughts in S frame generate a personal evaluation, whereas we thoughts generate a group evaluation. The solution concept in

the model roughly states that the cooperative option is chosen by a player in S if it is the best (against the mix of the co-player’s generated by his varying frame - p.21 -) in group evaluation and not worse than the other option in personal evaluation.

The S-frame intuition of 1997 unpublished paper, however, disappeared in subsequent pieces of work, but also the hypothesis that agent can ‘vacillate’ between the two frames does not appear in the most recent Bacharach’s effort: the book.

Later on, in developing the VFT Bacharach faces the issue of integrability of frames. He says that normally frames are integrable: “It is easy to integrate frames which consist of classifiers such as shape, colour and position: we can easily see a mark as a triangle, as a blue triangle, as a blue triangle on the left, . . . on the other hand. . . a person can see the marks as letters and as geometric shapes, but not at the same time - you can’t integrate these two perceptions” (2001, p.6). There exist frames, then, that are non-integrable. ‘I’ and ‘we’ frames appear non integrable in Bacharach’s words, and when this happens, “the agent may find herself vacillating between the judgments that she should do”(ib.). In spite of these attempts, then, ‘double-cross’ threat has not been enclosed in Bacharach’s analytical framework.

The second unsolved problem is the endogenization of ω . Bacharach tries to endogenize ω , because he sees that the fact that ω is exogenous represents a lacuna in his theory. We may see a clear evolution in his thoughts along the years: in 1997 and 1999

papers he sees ω as a function of the gain from cooperation and the harmony of interests: “To endogenize ω , and other feature of ω , one must show that the payoffs and other constitutive features of the basic game make collective identity salient or otherwise tend to induce team-thinking. The laboratory evidence is promising, as it suggests that group identification may be induced by the ‘common problem’ mechanism. In addition, it is plausible that ω may be an increasing function of certain quantitative features of the payoff structure, such as ‘scope for cooperation’ and ‘harmony of interest’” (1999, p.144). In 2001 he turns to other two features: “We need a link from the game parameters to the onset of group identification. Two of the classic favouring conditions are commonness of interests and commonness of predicament” (2001 b, p.8). In his book he goes deeper into the matter, as we have seen, by stating the interdependence hypothesis. But he does not complete the work: “It may also be that there is a positive relationship between salience and effectiveness: when a feature tending to promote self-identity is highly salient, then if and when it is noticed it is also highly effective. These are empirical speculations; their investigation will be an important part of the future development of the theory of group action” (2006, p. 87).

A step forward on this topic has been made by Tan and Zizzo (Tan, J. and D. Zizzo 2008): in their paper there is an attempt to investigate the relationship between harmony of interests (‘game harmony’ for them), group identification and cooperation. They claim that game harmony is a good

measure of the extent of cooperation or conflict in games. However, in experiments and in real life, we often observe what they call excess conflict or cooperation relative to theoretical predictions. In their framework, it is group identification that can explain excess in-group cooperation and excess out-group conflict. They also believe that the so called ‘perceived harmony’ can capture the effect of team reasoning on cooperation. This is a theoretical construct and it has not been tested yet. At the same time ‘perceived harmony’ cannot be inferred by the payoffs of the game, because it is a sort of ‘experimental’ measure (similarity index in experiments - how subjects think a game is similar to a pure coordination game or to a zero-sum game - is a proxy for perceived harmony). Anyway, the simple game harmony (not the perceived one) is the best existent proxy for what Bacharach has called ‘the harmony of interest’, and it is entirely derived from the payoffs of the game. Game harmony, defined as “a generic property describing how harmonious or disharmonious the interests of players are, as embodied in the payoffs” (Tan and Zizzo 2008, p. 3), is based on the correlation coefficient between payoff pairs. This measure can be a solution of Bacharach’s problem of endogenization of ω .

5. Conclusion

In this article I offered a review of theories of team reasoning. In particular I have analysed Bacharach’s theory of we-thinking. I focused on his attempt to for-

malize we-thinking in terms of Variable Frame theory. I found that two main problems arise trying to do it. One is linked to the endogenization of ω , the probability that a person may group-identify in a determinate situation. The other one is the impossibility, in Bacharach's framework, to use more than one frame at the same time. This problem prevents Bacharach from representing his 'double-crossing' intuition in the Prisoner Dilemma game, because it requires an agent to have I-frame in mind when he is we-thinking.

References

- BACHARACH, M. (1993): "Variable Universe Games," in *Frontiers of Game Theory*, ed. by K. Binmore, A. Kirman, and P. Tani. Massachusetts: MIT Press.
- BACHARACH, M. (1995): "Co-Operating without Communicating," London.
- BACHARACH, M. (1997): "'We' Equilibria: A Variable Frame Theory of Cooperation," Oxford: Institute of Economics and Statistics, University of Oxford, 30.
- BACHARACH, M. (1999): "Interactive Team Reasoning: A Contribution to the Theory of Cooperation," *Research in Economics*, 53, 30.
- BACHARACH, M. (2001): "Framing and Cognition in Economics: The Bad News and the Good," ISER Workshop, *Cognitive Processes in Economics*.
- BACHARACH, M., N. GOLD, and R. SUGDEN (2006): *Beyond Individual Choice*. Princeton University Press.
- GILBERT, M. (1989): *On Social Facts*. Routledge.
- GOLD, N., and R. SUGDEN (2008): "Theories of Team Agency," in *Rationality and Commitment*, ed. by P. Di Fabienne, and S. H.: Oxford University Press.
- HODGSON, D. H. (1967): *Consequences of Utilitarianism*. Oxford: Clarendon Press.
- HOLLIS, M. (1998): *Trust within Reason*. Cambridge: Cambridge University Press.
- HUME, D. ([1739] 1978): *A Treatise of Human Nature*. Oxford: Oxford University Press.
- HURLEY, S. (1989): *Natural Reasons*. Oxford: Oxford University Press.
- REGAN, D. (1980): *Utilitarianism and Cooperation*. Oxford: Clarendon Press.
- SUGDEN, R. (1993): "Thinking as a Team: Toward an Explanation of Nonselfish Behavior," *Social Philosophy and Policy*, 10, 69-89.
- SUGDEN, R. (2000): "Team Preferences," *Economics and Philosophy*, 16, 175-204.
- SUGDEN, R. (2003): "The Logic of Team Reasoning," *Philosophical explorations*, 16, 165-181.
- SUGDEN, R. (2005): "Fellow-Feeling," in *Economics and Social Interactions*, ed. by B. Gui, and R. Sugden: Cambridge University Press.
- TAN, J., and D. ZIZZO (2008): "Groups, Cooperation and Conflict in Games," *The Journal of Socio-Economics*, 37, 1-17.
- TUOMELA, R. (1995): *The Importance of Us: A Philosophical Study of Basic Social Notions*. Stanford University Press.

ZIZZO, D., and J. TAN (2003): "Game Harmony as a Predictor of Cooperation in 2 X 2 Games: An Experimental Study," Oxford: Department of Economics, University of Oxford.

ZIZZO, D. (2004): "Positive Harmony Transformations and Equilibrium Selection in Two-Player Games," Oxford: Department of Economics, University of Oxford.

On the Demand and Effectiveness of Reward in Voluntary Contribution Mechanisms. A Note

.....
Daniela Grieco
Università Bocconi
(Italia)

1. Introduction

Peers' punishment of free riders has been shown to represent a powerful instrument to induce cooperation in public good games (Fehr & Gaechter, 2000) and common pool resources (Ostrom, Walker & Gardner, 1992). Interestingly, the symmetric instrument of rewarding high contributors has exhibited significantly lower demand and efficacy in promoting cooperation. This result is robust across experimental settings (Dickinson, 2002; Andreoni, Harbaugh & Vesterlund, 2003; Walker & Halloran, 2004; Sefton, Shupp & Walker, 2006; Vyrastekova & Van Soest, 2008). Experimental subjects emerge - in general - not to elicit the option of rewarding cooperators as a device to drive peers' behaviour in favour of the group, neither altruistic behaviours are significantly stimulated by the possibility of being rewarded. As Offerman (2002) points out, "hurting hurts more than helping helps."

This note aims at discussing the reasons why individuals - in situations of anonymous interaction with strangers - do not decide to remunerate pro-social behaviours, whereas

sanctioning anti-social ones is quite commonly used and efficient in its impact on contributions levels. This result is suboptimal because rewards present both monetary and non-monetary benefits. Like sanctions, rewards are costly; however, they represent a transfer from the rewarder to the rewardee, whereas sanctions are costly for both the punisher and the punishee. Furthermore, and crucially, rewards do not induce in the recipient negative feelings like the anger and the desire of revenge that might undermine the cohesion of a group in case of punishment. An institution where rewards are fruitfully used as instruments to promote cooperation could represent a form of "virtuous" institution that rules out any constrictions and threats and where interactions occur in a more helpful environment. The absence of perverse effects and its nature of transfer are the reasons why an efficient use of rewards could represent a significant improvement in the devices put at work to prevent the decay in contributions. We believe that a better comprehension of this phenomenon could help in detecting the conditions where rewards could play a significant role.

In what follows, we address the issue of understanding the reasons behind the low demand and the ineffectiveness of reward resulting in economic experiments. The economic perspective is enlarged with insights from cognitive and social psychology. In particular, we analyse how rewards are perceived from a cognitive point of view and how the possibility of being rewarded does affect the motivations driving altruistic behaviours. Two distinct explanations for low demand

and ineffectiveness are discussed. On the one hand, rewarding is not elicited by subjects because it does not enable to yell out the negative emotions that instead can be expressed by assigning a sanction. On the other hand, rewards are ineffective in promoting cooperation because they are received by subjects holding already an intrinsic motivation to contribute: the extrinsic motivation built up by a reward or a sanction could be differently effective if an intrinsic motivation is already at work.

2. Economic Experiments On Reward

Experimental studies show the emergence of cooperation under conditions that strongly favour egoism. In real context of repeated, *non-anonymous* interactions, unlike any other species, humans cooperate with non-kin in large groups: this behaviour is puzzling from an evolutionary perspective because cooperators face individual costs to confer benefits on unrelated group members. In laboratory experiments, people interacting *anonymously* elicit positive contributions levels that decline as decisions are repeated: when given the opportunity to prevent the decay in contributions by costly punishing non-cooperators, subjects *do punish* even in one-shot interactions. Ethnographic data suggest that such altruistic punishment helps to sustain cooperation in human societies (Boyd, Gintis, Bowles & Richerson, 2003).

A bunch of studies focuses on the effectiveness of reward as an alternative or addi-

tional device to promote cooperation and prevent the decay in contributions. In two-stage games, a reward

is an increase in payoffs that a second-stage player imposes on a first-stage player, whereas a punishment is a decrease in payoff at the second stage. Under peculiar conditions, providing subjects with both the opportunity to reward *and* to punish results in higher efficiency, although more often this does not occur. Andreoni, Harbaugh and Vesterlund (2003) examine demands for rewards and punishments in a proposer-responder game. The proposer first makes an offer to split a fixed-sized pie; the responder is given or is *not* given a costly option of increasing or decreasing the proposer's payoff. The results show substantial demands for both punishments and rewards: however, rewards *alone* have little influence on cooperation, whereas punishments have some. When rewards and sanctions are combined, the effect on cooperation is relevant, suggesting that rewards and punishments are complements in producing cooperation. More specifically, offers exceed the Nash equilibrium level in all three treatments and, interestingly, this effect is *stronger in the rewards treatment* than in the sanctions treatment. However, offers are highest in the treatment where both instruments are available to the responder.

Sefton, Shupp and Walker (2003) contrast the institutions of reward and punishment with respect to their impact on cooperation and efficiency in the context of voluntary contributions mechanism with linear payoffs. In the reward treatment, contributions decrease over time reaching a level below the one ob-

served in the absence of opportunities to reward: the opportunity to reward results insufficient to sustain cooperation; in contrast, sanctions succeed in sustaining public goods provision at a level above the one observed in the absence of sanctioning opportunities, and so sanctioning appears to be a more effective mechanism for sustaining contributions. However, opportunities to sanction initially result in a loss of efficiency, as the direct costs associated with sanctioning prevail over the effect of increased contributions. The authors also observe differences in the dynamic patterns of rewarding and punishing behaviour: whereas subjects initially use rewards more frequently than sanctions, over time the use of rewards declines at a faster rate than the use of sanctions, so that in later rounds rewards are used less frequently than sanctions. These results capture the complexities involved in using a reward system for sustaining cooperation: “one might argue that a successful reward system requires continued use of rewards, and those rewards must be in the form of transfers from those allocating less to the public good to those allocating more. In our experiments, however, it was those subjects who allocated relatively more to the group account who tended to give more rewards” (*ibidem*). The significant decay in rewards across decision rounds suggests that *groups may have difficulty in maintaining a rewards system* and that subjects seemed to lack a clear focal point or consensus in regard to where rewards should be targeted. Moreover, rewards seem to work only with a critical mass of altruistic individuals that contribute in the first stage and then reward who contributed in

the second stage: however, if this is true, there would be no need of building an extrinsic motivation by means of rewarding or punishing because it would be effective for people who already behaved altruistic. In a similar flavour, Dickinson (2001) studies teams of four members in a laboratory environment where giving effort towards a “team goal” is captured by eliciting voluntary contributions towards the provision of a public good. The efficiency-improving properties of four distinct environments is tested: monetary prizes given to high contributors versus monetary fines assessed to low contributors, where being high or low contributor is defined first in terms of absolute contributions and then in terms of contributions relative to abilities (called “handicapping”). The use of carrot (positive) and stick (negative) incentives are methods of increasing effort among members of work teams. The results show that both carrot and stick can significantly increase efficiency (i.e., contributions): more specifically, handicapped incentives promise the highest efficiency levels and, when handicapping is not used, *penalties are more effective than prizes*.

Walker and Halloran (2004)’s study involves a one-shot, voluntary contribution mechanism and finds that, whereas some subjects are willing to reward and sanction others at a private cost, *the opportunity to reward or sanction is ineffective in facilitating cooperation* relative to previous experiments in which repeated game environment is employed. Andreoni, Harbaugh and Vesterlund (2003)’s and Sefton, Shupp and Walker (2003)’s experiments present anonymous, strangers treatments: however, a reputation effect about

peers' attitude towards cooperation, or at least a mechanism of learning about the rules of the game, seems to emerge and strengthen the role of punishment and, in a lower extent, of reward. This effect disappears in a one-shot game like Walker and Halloran (2004)'s.

Vyrastekova and Van Soest (2007)'s work studies the effectiveness of costly rewards in mitigating excess extraction in a standard Common Pool Resource game experiment. Two treatments are implemented: in the first, rewards are a pure transfer from one player to the other ("transfer reward"); in the second, the benefits of receiving a reward are higher than the cost of providing it ("net positive reward")¹. In this perspective, the authors suggest an explanation for Andreoni, Harbaugh and Vesterlund (2003)'s results on reward effectiveness: the difference with the ineffectiveness emphasized by other experiments' results might lie in the *parameterization*. In fact, Andreoni, Harbaugh and Vesterlund (2003)'s study uses either 1:5 sanctions, or 1:5 rewards, or both. Vyrastekova and Van Soest find that, when the benefits of receiving a reward exceed the cost of providing it, the use of extraction tokens in the CPR game is significantly closer to the social optimum than in case of transfer rewards. This difference in effectiveness arises due to the difference in profitability of receiving rewards (and consequently in the opportunity costs of not receiving them), and due to differences in who rewards whom. When there are positive net gains from exchanging reward tokens, estab-

lishing bilateral exchange by means of reward is in the interest of both the players. As subjects who cooperated have no incentives to send reward tokens to free-riders, bilateral exchange is established among subjects using the same number of extraction tokens. This represents for free-riders an incentive to decrease the number of extraction tokens used to attract rewards. Such considerations are absent when there are zero net gains from exchanging rewards, as in the case of the transfer treatment. This kind of analysis is important because in many situations, a rewardee's evaluation of a reward is unlikely to be equal to the costs incurred by the rewarder. In case of rewards taking the form of monetary transactions, this can be due to subjects' differences in income or wealth; in case of rewards taking the form of the exchange of goods or services, marginal utilities may differ considerably between individuals.

Vyrastekova and Van Soest (2007)'s contribution is very relevant in the topic and sheds light on the reasons why sanctions are generally more effective than rewards. However, it lacks in investigating explicitly the difference between sanctions and rewards. Although the role of the parameterization and the prize-to-fee ratio are significant, a change in the impact ratio has been found to substantially affect the effectiveness of sanctions (Carpenter, 2006; Egas & Riedl, 2005) as well. As Casari (2007) shows, sanctions are sensitive to parameterization. Rewards behave coherently in the same directions: to prove lower effectiveness of the rewards, what would matter is a relatively higher sensitivity

to parameterization that has not been documented yet.

3. The Psychological Literature On Positive Reinforcement

Psychological studies strongly differ from economics in predicting the role of rewards versus sanctions. The psychological literature conceives rewards (sanctions) as positive (negative) reinforcements of actions. A huge amount of studies within the operant tradition (since the seminal work by Skinner, 1953 on) has documented that *extrinsic rewards do shape behaviour significantly*.

Psychological studies on punishment and reward are concerned with adaptive learning and the contrast between intrinsic and extrinsic motivation. Altruistic and pro-social behaviour are forms of conduct that derive from intrinsic motivations but that can be stimulated by means of extrinsic devices like reward or punishment. Four types of outcomes can occur in operant research (Burgess and Bushell, 1969): (1) positive reinforcement, i.e. the addition of an appetitive stimulus to a situation; (2) negative reinforcement, i.e. the removal of an aversive stimulus from a situation; (3) positive punishment, i.e. the addition of an aversive stimulus to a situation; and (4) negative punishment, i.e. the removal of an appetitive stimulus from a situation. The two types of reinforcement can be grouped under the general label of “reward,” while the two types of punishment can be grouped under the general label of “cost” (Gray & Tallman, 1987). By 1971, hundreds of studies within the oper-

ant tradition had established that extrinsic rewards can control behaviour. In fact, when administered closely subsequent to a behaviour, rewards were reliably found to increase the likelihood that the behaviour would be emitted again, an effect that persisted as long as the reward contingency was operative. When rewards were terminated, the likelihood that the behaviour would be emitted eventually returned to the pre-reward baseline. This general finding led to the widespread advocacy of rewards as a motivational strategy, and behaviour-change programs based heavily on the use of rewards were introduced into a variety of applied settings (Deci, Koestner & Ryan, 1999).

A wide range of studies has explored the effects of various types of rewards like tangible rewards such as money (Deci, 1971; Deci, 1972a, Deci, 1972b), positive feedback or verbal rewards. Additional studies explored the effect of external effects such as classroom climates (Deci, Schwartz, Sheinman & Ryan, 1981) and the combined effect of reward and interpersonal contexts (Ryan, Mims & Koestner, 1983). Interpersonal contexts represent the social ambience of settings like homes, classrooms, work groups and play a role in the extent to which people within the context feel pressured to behave in a particular way. When studied in laboratory experiments, the interpersonal climate is usually instantiated as the interpersonal style used by the experimenter to administer rewards (Ryan, 1982). An important work by Miller, Butler and McMartin (1969) on social power emphasizes the role of reward and punishment as a form of social control whose strength and effect de-

rive from the ability to deliver reward and punishment. The authors remark an asymmetry between reward and punishment based on two arguments. First, considerations on notions like exchange (e.g. Blau, 1964), reciprocity (Gouldner, 1960) and equity (Adams, 1964) suggest that reward and punishment implemented by high level of power subjects should imply distinct (instead of similar) outcomes due to the fact that high power subjects that use reward should reach their goals because their use of rewards is perceived as altruistic and generous. On the contrary, high power subjects that exercise their power by means of sanctions could have difficulties in getting their ends and, if they do, “in their case their dessert are aversive stimuli” (Miller, Butler & McMartin, 1969). Furthermore, reward and punishment differ in their effect on the learning and maintenance of human behaviour (Estes, 1944; Skinner, 1953): mild punishment seems to be less permanent in its effects, is more likely to produce unwanted emotional behaviour than does reward and the behaviour that it elicits may be hard to specify. However, when subject face stochastic punishment or reward, subjective inferences on the perceived probability of being punished or rewarded play a role. The huge amount of work generated from Tversky and Kahneman and their colleagues (e.g., Kahneman et al., 1982), suggests that a given probability of punishment may be a greater deterrent against crime than an equal probability of rewards for choosing a non-criminal activity. This follows since the probability of negative punishment for a non-criminal activity would be greater, given that success at non-criminal

activity was a rare event; in such situations the stick may be more powerful than the carrot.

4. Discussion

The insights deriving from the experimental studies and the psychological investigations reported above shed light on the discrepancy in demand and effectiveness between rewards and sanctions. In this section, the possible explanations of this phenomenon are presented and discussed.

An explanation for low demand: Motivational factors. Psychological studies focus on the fact that an extrinsic device such as reward might undermine intrinsic motivation to cooperate. As emphasized by Charness (1999), recent models of non-pecuniary motives to pro-social behaviour can be classified as either altruism-based, equity-based, or reciprocity-based. One explanation is altruism, where people care not only about their own material well-being but also about the material well-being of others: an agent is assumed to assign a constant weight to the welfare of every other individual. Another approach expands the altruism principle by incorporating distributive concerns: these models, known as equity or fairness models, one’s regard for another person’s monetary well-being depends on the other person’s monetary payoff relative to one’s own⁴. Finally, reciprocity theories expand the principle of altruism in a different direction, asserting that attention for someone else’s payoff depends on how much altruist the other is perceived to be: the prin-

ciple of reciprocal altruism emphasizes the relationship between one's action and one's beliefs about the intentions of the other agents⁵.

4 Current models in this category are Bolton and Ockenfels (1997) and Fehr and Schmidt (1997). 5 The leading model in this category is Rabin (1993), while Levine (1996) offers an alternative formulation.

These issues of altruism and selfishness drive some crucial questions on our evolutionary origins, our social relations, and the organization of society. Experimental evidence indicates that, although agents are self-interested and maximize their earnings, human altruism is a powerful force. However, there is heterogeneity across individuals: economic agents are not generally selfish but might belong to further types, i.e. be altruists or reciprocators. In experiments, different player "types" seem to emerge: (1) some subjects are standard selfish maximizers, confirming that the *homo oeconomicus* exists; (2) other subjects act altruistically, as if they were driven by other-regarding preferences; (3) a third group of players are willing to non-strategically punish and/or reward others, i.e. are driven by some form of "strong reciprocity", based on evaluation of the opponent's intentions (Camerer & Fehr, 2006). It is this interaction between altruist and selfish individuals that drives human cooperation (Fehr & Fischbacher, 2003). The fact of belonging to a specific type is related to the presence or absence of intrinsic motivations (of being altruistic): only altruists have an intrinsic motivation that moves contribution to a public good. Following the psychological debate on the role of positive, extrinsic reinforcement on

intrinsic motivations, we argue that a possible explanation of the lower effectiveness of reward relies on the fact that rewarding an altruistic behavior undermines the presence of an intrinsic motivation to altruism. If people are altruist and believe that their pro-social behavior is interiorly grounded and this should be true also for peers, for instance due to the importance of a common project, the presence of a device that should stimulate pro-social behavior could eliminate this feeling towards the project. If a reward is directed to pay off altruists, they would probably argue that there should be some reason not to be altruistic. Any spontaneous effort is ruled out. On the contrary, sanctions may build an (extrinsic) motivation of behaving altruistically for individuals who have no intrinsic motivation and, in absence of sanctions, would free-ride. In this perspective, it would be interesting to explore more specifically whether the ineffectiveness of reward (contrasted to sanction) is due to the fact that it undermines the intrinsic motivation of a purely altruistic choice. Hence, a comparison between altruists and selfish/reciprocators in their reactions of altruists to reciprocators to rewards should be carried out: only altruists' behavior should be affected by such a reduction in intrinsic motivation.

An explanation for ineffectiveness: The role of emotions. A relevant source of asymmetry between reward and punishment lies in the type and strength of emotions that originate them and that they provoke in targets. Fehr and Rockenbach (2003) point out that in human societies, social order and cooperation rely on

both the use of rewards and sanctions, “which ensures the compliance of self-

interested actors, and on the presence of people willing to perform altruistic acts”. Their experiment present a setup based on dyadic interaction in trust game. In each round of the experiment two mutually anonymous subjects (an investor and a trustee) are involved. First, the investor has the chance of choosing a costly trusting action; then the trustee is informed about the investor’s action and can honour the investor’s trust by taking a costly cooperative action. If the investor chooses a trusting action and the trustee responds cooperatively, both players increase their monetary payoff, but the trustee has the option of not honouring the investor’s trust, saving the costs of cooperation but causing the investor to be worse off. Consequently, if the investor expects that the trustee will not honour his trust, he will not choose a costly trusting action at the beginning of the game. Therefore, the subjects face a dilemma because trust and cooperation are advantageous for both subjects but the trustee faces the temptation not to cooperate and, therefore, the investor is tempted not to trust. The main results in Fehr and Rockenbach’s experiments is that sanctions’ effectiveness relies on the motivation that has previously driven the sanction: whereas altruistically motivated sanctions implemented for the benefit of the group enhance cooperative behaviour, sanctions that are imposed to enforce an unfair distribution of resources have the opposite effect. Previous research on public goods experiments has shown that altruistic punishment is a highly effective means of enforcing

cooperation. Sanctions in the public good context enhance cooperation only when the punishment of free-riders is an altruistic act that is considered as morally legitimate, whereas rewards should not present the same need of being legitimated.

Offerman (2002)’s experiments present simple proposer-responder games that demonstrate that responders are willing to depart from own-earnings maximization by rewarding more generous proposers or sanctioning less generous proposers. The main hypothesis tested in this paper is that negative intentions provoke stronger reciprocal responses than positive intentions. The potential asymmetric effect of positive and negative intentionality is assessed in the “hot response game”. In this game the first mover makes a choice between a helpful and a hurtful choice where the helpful choice increases the payoff of the first mover by less guilders than the hurtful choice and increases the payoff of the second mover by the same number of guilders. The second mover observes the choice of the first mover, before choosing between a cool, remunerative response and a hot, reciprocal response. People easily feel insulted after a hurtful choice. In accordance with this conjecture, subjects experience stronger negative emotions when a hurtful choice is intentional than when it is unintentional. In addition, they show more restraint in responding to their emotion when the hurtful choice was unintentional. Both an increase in the negative emotion experienced and a decrease in the restraint of responding to the emotion explain the strong effect of negative intentionality on the probability of reciprocation. Negative intentionality matters

considerably more than positive intentionality.

Over the years, numerous studies of law and economics exposed the effect of legal sanctions on individual behaviour, whereas neuroeconomics studies have been contributing to indicate that the desire to punish is often strong and non-strategic, and that it is a largely automatic response triggered by emotional forces (see Singer and Fehr, 2005). Among others, the positron emission tomography has been used to examine the neural basis for altruistic punishment of defectors in an economic exchange in an experiment where subjects could punish defection either symbolically or effectively. Symbolic punishment did not reduce the defector's economic payoff, whereas effective punishment did reduce the payoff. Effective punishment, as compared with symbolic punishment, activated the dorsal striatum, which has been implicated in the processing of rewards that accrue as a result of goal-directed actions. Moreover, subjects with stronger activations in the dorsal striatum were willing to incur greater costs in order to punish. These findings support the hypothesis that people derive satisfaction from punishing norm violations and that the activation in the dorsal striatum reflects the anticipated satisfaction from punishing defectors. This result is confirmed by Casari and Luini (2008)'s finding that punishment is expressive, and individuals do not consider sanctions inflicted by peers as a substitute of their own sanction, although sanctioning is costly.

They seem to accrue an intrinsic pleasure to punish by themselves, and the possibility to

set out negative emotions by punishing deviators should play a role.

In this perspective, it would be interesting to explore whether the effect of the negative emotions in strengthening the efficacy of sanctions compared to rewards could be removed by implementing virtuous institutions where punishment cannot be used to implement anti-social behaviours like targeted or blind revenge that grounds on negative emotions. This idea is confirmed by Haigner, Kocher and Sutter (2006), who analyze an experimental public goods game in which group members can endogenously determine whether they want to supplement a standard voluntary contribution mechanism with the possibility of rewarding or punishing other group members. They find a large and positive effect of endogenous institutional choice on the level of cooperation in comparison to exogenously implemented institutions: democratic participation rights enhance cooperation in groups. Moreover, with endogenous choice, groups typically vote for the reward option.

References

- Adams, J.S. (1965). Inequities in social changes, *Advances in Experimental Social Psychology*.
- Antoci, A. and Zarri, L. (2008). On Punishing Non-Punishers. When (Not So) Nice Guys Deserve the Stick, not the Carrot, Science24.com
- Andreoni, J., Harbaugh, W. and Vesterlund, L. (2003), The Carrot or the Stick: Re-

- wards, Punishments, and Cooperation, *American Economic Review*, 93 (3), 893-902. 15
- Blau, P. M. (1964), *Exchange and Power in Social Life*. John Wiley & Sons, New York.
- Boyd, Gintis,
- Bowles and Richerson, (2003), The evolution of altruistic punishment, *PNAS*, March 18, 100 (6), 3531–3535.
- Burgess, R.L. and Bushell, D. (1969), *Behavioral Sociology: The Experimental Analysis of Social Process*. Columbia University Press.
- Camerer, C. and Fehr, E. (2006). When does “Economic man” dominate social behaviour?, *Science*, 311, 5757, 47-52.
- Casari, M. (2005). On the design of peer punishment experiments, *Experimental Economics*, 8, 107-115.
- Charness, G. and Haruvi, E. (2002). Altruism, equity, and reciprocity in a gift-change experiment: An encompassing approach, *Games and Economic Behavior*, 40(2), 203-231.
- Dawes, R.M. (1980). Social dilemmas, *Annual Review of Psychology*, 31, 69-93.
- Deci, E. L. (1971), Effects of externally mediated rewards on intrinsic motivation, *Journal of Personality and Social Psychology*, 18, 105-115.
- Deci, E. L. (1972a). Effects of contingent and non-contingent rewards and controls on intrinsic motivation, *Organizational Behavior and Human Performance*, 8, 217-229. 16
- Deci, E. L. (1972b). Intrinsic motivation, extrinsic reinforcement, and inequity, *Journal of Personality and Social Psychology*, 22, 113-120.
- Deci, E. L., Koestner, R. and Ryan, R.M. (1999), A meta-analytic review of experiments examining the effect of extrinsic rewards on extrinsic motivation, *Psychological Bulletin*, 125(6), 627- 668.
- Dickinson, D. (2000), The Carrot vs. the Stick in Work Team Motivation, Working Paper ERI 2000-06 Utah State University.
- Fehr, E. and Fischbacher, U. (2003), The nature of human altruism, *Nature*, 425, 785-91.
- Fehr, E. and S. Gächter (2000), Fairness and retaliation: The Economics of reciprocity, *Journal of Economic Perspectives*, 14, 159-81.
- Fehr, E. and Rockenbach, B. (2003), Detrimental effects of sanctions on human altruism, *Nature*, 137-140.
- Fehr, E. and Schmidt, K. (1999), A theory of fairness, competition, and cooperation, *Quarterly Journal of Economics*, 114, pp. 817–868
- Flache, A. (1996), *The Double Edge of Networks. An Analysis of the Effect of Informal Networks on Cooperation in Social Dilemmas*, Amsterdam: Thesis Publishers
- Gouldner, A.W. (1960), The norm of reciprocity: A preliminary statement, *American Sociologic Review*, 25(2), 161-178. 17
- Gray, L.N. and Tallman, I. (1987), Theories of choice: Contingent reward and punishment applications, *Social Psychology Quarterly*, 50 (1), 16-23.
- Haigner, S., Kocher, M. and Sutter, M. (2006), Choosing the Stick or the Carrot? Endogenous Institutional Choice in Social Dilemma Situations, *CESPR Discussion Paper*.

- Kahneman, D., J. Knetsch and R. Thaler (1986)., Fairness and the assumptions of economics, *Journal of Business*, 59, pp. S285–300.
- Miller, N. Butler, D.C. and McMartin, J.A. (1969) ,The ineffectiveness of punishment power in group interaction, *Sociometry*, 32 (1), 24-42.
- Molm, L.D. (1997), *Coercive Power in Social Exchange*.
- Offerman, T. (2002), Hurting Hurts more than Helping Helps, *European Economic Review*, 46(8), 1423-1437.
- Ostrom, E., J. Walker and R. Gardner (1992). „Covenants with and without a sword: Self- governance is possible’, *American Political Science Review*, 86, 404-417.
- Ryan, R. M. (1982)., Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory, *Journal of Personality and Social Psychology*, 43, 450-461. 18
- Ryan, R. M., Mims, V. and Koestner, R. (1983), Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory, *Journal of Personality and Social Psychology*, 45, 736-750.
- Sefton, M., Shupp, R. and Walker, J. (2003)., The effect of rewards and sanctions in provision of public goods’, Indiana University Working Paper.
- Sell, J., Chen, Z., Hunter-Holmes, P. and Johansson, A. C. (2002)., A Cross-Cultural Comparison of Public Good and Resource Good Settings, *Social Psychology Quarterly*, 65 (3), 285-297.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
- Son, J. and Sell, Y. (1995)., Comparing public goods with common pool resources: Three experiments, *Social Psychology Quarterly*, 60 (2), 118-137.
- Vyrastekova, J. and Van Soest, J. (2007)., On the (in)effectiveness of rewards in sustaining cooperation, *Experimental Economics*, 11, 53-65.
- Walker, J.M. and Halloran, M. (2004)., Rewards and sanctions and the provision of public goods in one-shot settings’, *Experimental Economics*, 7, 235–247.

Maria Carolina Foi, *La giurisdizione delle scene. I drammi politici di Schiller* (Macerata: Quodlibet, 2013).

La monografia di Maria Carolina Foi è il frutto di un lavoro di indagine decennale, inizialmente avviato sotto l'egida della Alexander von Humboldt-Stiftung e proseguito non in ultimo grazie al sostegno del Deutscher Akademischer Austausch-Dienst (D.A.A.D.). Nel volume vengono ripresi e ulteriormente sviluppati alcuni saggi già pubblicati tra il 2001 e il 2011, dedicati ad aspetti parziali del grande tema in oggetto. Ma non solo. Altri contributi inediti arricchiscono l'insieme donandovi unità e coesione.

Il tema in quanto tale si inserisce in una felice congiuntura: da un lato, la più o meno recente ripresa di interesse per lo Schiller politico a partire dagli studi di Hans-Jürgen Schings (1996, 2012), Jeffrey L. High (2004), Walter Müller-Seidel (2009) e Susanne Aigner (2012), dall'altro la vera e propria scoperta dello Schiller giurista, inizialmente e non a caso monopolio di storici del diritto come Klaus Lüderssen (2005) e Matthias Tresselt (2009), ma negli ultimi anni rivendicato come legittimo oggetto di studio dalla germanista Yvonne Nilges (2012). Nel primo caso, la nuova stagione della ricerca può essere vista anche come conseguenza del mutato clima ideologico e dello sgretolamento di visioni del mondo pregiudiziali a una serena valutazione dell'impegno – alcuni vorrebbero: del disimpegno – schilleriano nei confronti della politica del suo tempo. Nel secondo, l'esplorazione della *terra inco-*

gnita del diritto prosegue sul sentiero indicato da Wolfgang Riedel a proposito della formazione medica del giovane Schiller (1985) e in seguito proficuamente percorso anche in Italia – basti pensare alla recentissima edizione italiana degli scritti giovanili di Schiller a cura di Giovanna Pinna (2012). Ora è la volta dell'istruzione giuridica che il giovane Schiller ricevette alla *Karlschule* di Stoccarda tra il 1774 e il 1775, prima cioè che uno scarso rendimento scolastico e quella che egli avvertiva essere una maggiore affinità della medicina con la poesia lo conducessero a cambiare indirizzo di studio. Certo, già nel 1972 Klaus Oettinger aveva richiamato l'attenzione sul *j'accuse* giuridico implicito nel notissimo racconto *Der Verbrecher aus Infamie* (1786), mirante a mettere in luce le falle di certo diritto penale e, per converso, a proporvi dei mutamenti che andassero a sostituire l'attenzione esclusiva al reato con una più ampia considerazione del reo. Le conseguenze di questo pionieristico contributo si erano però lasciate attendere, e con esse una rivoluzione negli studi analoga a quella che invece Riedel sarebbe poi riuscito a provocare con la sua monografia.

Evidentemente, il volume della Foi si caratterizza per tempestività storiografica su due fronti. Sul primo, smentisce con nettezza le ormai obsolete letture di uno Schiller “borghese”, rinunciatario e provinciale, colpevole di un codardo escapismo nei confronti non solo e non tanto della realtà politica del suo tempo, ma anche, e forse soprattutto, della propria produzione giovanile. Sul secondo, corregge le unilateralità metodologiche di approcci giusletterari “strumentali”,

volti cioè semplicemente a rinvenire ed estrapolare le dottrine giuridiche disseminate nel corso dell'opera poetica e drammaturgica, senza alcuna considerazione per il medio letterario a cui queste sono affidate. È forse dovuta a questa presa di posizione metodologica l'omissione di qualsiasi riferimento alla formazione giuridica di Schiller, che se da un lato avrebbe garantito completezza al paragrafo iniziale intitolato *Schiller giurista?* (pp. 11-21), dall'altro avrebbe sicuramente comportato quella, «peraltro auspicabilissima, verifica storico-sistematica dei contatti dell'autore con la cultura giuridica a lui contemporanea» che la Foi esclude programmaticamente dal suo ambito di indagine (p. 30) – e che, sia lecito qui aggiungerlo per completezza, viene comunque intrapresa dalla Nilges nella sua monografia.

Il volume è costituito da sei capitoli disposti secondo un criterio legato ai generi letterari: i primi due capitoli, rispettivamente intitolati *Metodo: un frammento di Schiller sul diritto naturale* (pp. 11-34) e *Il diritto alla rivoluzione. Schiller e Erhard* (pp. 35-74), tematizzano due saggi, probabilmente risalenti allo stesso torno di anni, vale a dire il frammento *Method* e la serie di lettere *Über die ästhetische Erziehung des Menschen*. I restanti capitoli si concentrano su alcuni fra i maggiori drammi schilleriani, compiuti e non, facendo leva di volta in volta su un problema giuridico-filosofico ben determinato: *I diritti delle donne e la filosofia dell'amore*. Don Carlos (pp. 75-107), *Un'irruzione del tempo nel dramma*. Maria Stuart (pp. 109-138), *I diritti dell'uomo e la dignità del femminile*. Wilhelm Tell (pp. 139-176), *L'usurpatore en bonne*

foi. *Questioni di legittimità nel Demetrius* (pp. 177-218).

Il primo capitolo funge sostanzialmente da introduzione metodologica e fa propria la contrapposizione del frammento *Method* tra l'approccio meramente teorico di diritto naturale, politica, morale ed estetica da un lato, e la caotica violenza della realtà dall'altro, al fine di rivendicare la maggiore appropriatezza di un'indagine che si rivolga all'opera letteraria rispettandone statuto e prerogative. Insomma, un'indagine che ricerchi il giuridico in Schiller non scindendo il contenuto dalla forma, ma sforzandosi di vedere il contenuto *anche* nella forma. Il secondo capitolo prende le mosse da un affondo biografico sull'amicizia tra Schiller e il medico giacobino nonché kantiano Johann Benjamin Erhard per procedere poi alla reinterpretazione della presunta fuga di Schiller nell'ideale come operazione di dislocazione tutt'altro che meramente difensiva rispetto all'urgenza della realtà. Una strategia offensiva, questa, che certo assume carattere obliquo, ma che non per questo fallisce il bersaglio. Evidentemente, i primi due capitoli precisano la posizione della Foi rispetto ai due fronti di cui sopra.

Il terzo capitolo è sostanzialmente retto da un doppio interesse, da un lato e sulla scia di quello che Schings ha fatto nel 1982 per *Die Räuber*, dalla volontà di interpretare il *Don Carlos* attraverso le categorie della *Theosophie des Julius*; dall'altro, e in singolare consonanza con le più recenti acquisizioni in materia di storiografia filosofica, dal tentativo di individuare una continuità nello sviluppo teorico e concettuale di Schiller oltre e al di là della

presunta svolta kantiana. Se Eboli e Posa mettono dunque alla prova la tenuta della metafisica dell'amore (peraltro già in decisa crisi al momento della stesura del dramma) con un amore che per vendetta si vende e un amore che per l'ideale dimentica l'amicizia, la contrapposizione, introdotta da Posa a proposito di Eboli e di Elisabeth, tra una virtù artificiale (Eboli) e una virtù inconsapevole (Elisabeth), sembra anticipare la richiesta postkantiana di una grazia che *dia l'impressione* di essere naturale. Che questa stessa contrapposizione sia poi il costrutto di un personaggio maschile che intende strumentalizzare entrambi i personaggi femminili, non fa che accrescere la complessità (e profondità!) di questa lettura, facendo risuonare un basso continuo di genere.

Il quarto e il quinto capitolo affrontano dottrine di più immediata appartenenza giuridica quali la trasparenza del procedimento processuale, la pena di morte, i diritti civili e i diritti umani. Ne risulta una lucida consapevolezza dei meccanismi di inclusione/esclusione e rimozione che determinano l'acquisizione dei moderni concetti di uguaglianza e giustizia: Maria Stuart muore come un qualsiasi condannato a morte (inclusione di un individuo d'eccezione nell'universale umano), dopo aver però denunciato i vizi di un processo intentato per motivi politici e del tutto dimentico della tutela dovuta anche a lei come persona (esclusione di questo stesso individuo dall'universale umano); i confederati del Rütli riescono nell'impresa di riconquista dei loro diritti originari solo grazie all'omicidio ai danni di Gessler che, a dispetto delle loro stesse convinzioni, non lascia

innocenti le loro mani (rimozione). Ma non è tutto, perché ancora una volta la riflessione deve muoversi su più piani: l'assassinio commesso da Tell mira infatti a compensare la colpa di quello potenziale del figlio, presentato non come violazione dei diritti del giusnaturalismo moderno né della loro sanzione rivoluzionaria, ma del diritto materno, in altre parole: della legge di natura declinata al femminile e impersonata nel dramma da Hedwig. Di nuovo, il basso continuo intona una melodia di genere.

Il sesto capitolo è la conclusione più adeguata per l'indagine proposta dalla Foi. Esso infatti riprende le fila del percorso intrapreso, ma non le conduce a un esito fermo, quasi chiusa sistematica di un itinerario volutamente presentato come aperto. Anzi, complice lo stato incompiuto del testo scelto come riferimento, vengono ribadite la modernità e la spregiudicatezza delle posizioni schilleriane, senza però volerle ricondurre a un preciso edificio politico, giuridico o filosofico. La legittimazione del potere diventa allora un problema non solo oggettivo, valutabile sulla base di dati concreti e regole codificate, ma anche soggettivo, constatabile sulla scorta della credibilità dello stesso titolare di questo potere, o aspirante tale, di fronte a se stesso e agli altri (Demetrius); il femminile si fa erede di un atteggiamento che dal Cinquecento non prende più la virtù, finta o autentica, delle dame alla corte di Filippo II, ma il cinico realismo della Firenze di Machiavelli (Marina); la democrazia rappresentativa, infine, si rivela teatro di una gestione viziata della politica, dominata da una maggioranza sconsiderata, peraltro ag-

girata dalle manovre messe in atto dietro le quinte (Parlamento di Cracovia).

Niente di più lontano, insomma, dalla narrazione di uno Schiller idealista presuntivamente impegnato a enunciare principi astratti e scarsamente interessato a verificarne poi l'effettiva realizzabilità. Al contrario, il gioco di specchi a cui viene ridotta la politica prova la lucidità e la lungimiranza dell'esercizio interpretativo compiuto da Schiller sul suo tempo. La Foi parla non a caso di contiguità con «celebri teorie novecentesche della sovranità», che si rovescia però subito in una sorta di «correzione preventiva di quelle teorie» (p. 114), e di «una sorprendente denuncia *ante litteram* dei limiti del parlamentarismo e delle sue procedure» (p. 195). Certo, la serietà metodologica propria di qualsiasi corretta storiografia, letteraria come filosofica, politica come giuridica, impone di precisare che non si tratta di «attribuire retrospettivamente al testo categorie filosofico-politiche novecentesche», quanto di «non dare per scontato ciò che semmai andrebbe visto come uno dei possibili effetti» dell'«operazione» schilleriana (p. 187). Ne va cioè di capire come Schiller riesca «attraverso il medium teatrale» a trattare temi tuttora di bruciante attualità teorica e, grazie a una «prestazione» della forma estetica [...] messa in luce attraverso l'indagine storica», arrivi in definitiva «fino al Novecento» (p. 34).

Evidentemente, il volume della Foi è ricco, suggestivo e convincente. Scritto per incidere sulla ricerca, può essere letto senza troppe difficoltà anche da studenti e dottorandi non ancora ben addentro alle questioni

trattate: la chiarezza espositiva e la trasparenza argomentativa sono un indiscutibile pregio del volume. Resta da sperare che la ricezione di un lavoro così valido e innovativo non venga pregiudicata dalla lingua in cui è scritto, o che, dovesse questo essere il caso, se ne possa presto avere un'edizione tedesca, magari aggiornata.

Laura Anna Macor

Barbara Scapolo, *Leggere Timore e Tremore di Kierkegaard* (Como – Pavia: Ibis, 2013).

Timore e tremore. *Lirica dialettica di Johannes de Silentio* (*Frygt og Bæven. Dialektisk Lyrik af Johannes de Silentio*, 1843) è forse l'opera più articolata e affascinante della cospicua e labirintica produzione kierkegaardiana; un'opera dal «*pathos* tremendo», scritta su «carta filigranata» – come lo stesso Kierkegaard annota nel suo *Diario* – e dallo stile suggestivo e malinconico.

Il lavoro ben scritto e ben strutturato di Barbara Scapolo si presenta non solo come valido percorso di lettura, una guida e un supporto preziosi per l'esplorazione di questo testo denso e complesso, ma anche come un'accurata e approfondita analisi corredata, tra l'altro, da un'«Appendice» con i passi dei *Papirer* che mostrano come Kierkegaard non solo abbia sostanzialmente abbandonato, nel prosieguo del suo itinerario di pensiero, la posizione iniziale di *Timore e tremore* marcando sempre più la differenza tra la religiosità tutta *terrena* di Abramo e quella tutta *ul-*

traterrena del cristianesimo, ma abbia anche privilegiato la prospettiva del Cavaliere dell'eterna rassegnazione a scapito di quella del Cavaliere della fede, che è invece centrale nella "lirica dialettica".

La lettura approntata da Scapolo, inoltre, è finemente contestualizzata in termini biografici, bibliografici e storico-culturali, dato che si sofferma, in maniera ragionevolmente propedeutica, sia sul Kierkegaard propugnatore di una *nuova* religiosità, di una spiritualità sincera e militante – vissuta in polemica con il "demoralizzante" idealismo tedesco, con l'inautenticità del proprio tempo e con la cristianità stabilita ("massificata e mondanizzata") della Copenaghen della *Golden Age* del XIX secolo –, sia sul Kierkegaard *pensatore privato* e teorico della categoria del Singolo che, livellato e spersonalizzato da un astratto, amorfo e imbarbarito "spirito del gregge" (la Folla, il numero), trova nel *coram deo* il coraggio di dire "io", di appropriarsi di sé, di ripetere o riprendere in maniera diversa ciò che ha perduto nell'*en masse*.

In costante riferimento a questi aspetti, il merito aggiuntivo che va riconosciuto al lavoro di Scapolo è l'opportunità affatto secondaria con cui si addentra nei meandri della strategia discorsiva di Kierkegaard per comprenderne le finalità meno evidenti. Nel primo capitolo (*L'esistenza singolare come problema*), infatti, la studiosa presenta un approfondimento del valore etico-performativo dell'uso della strategia comunicativa estetico-pseudonima in *Timore e tremore* che è strettamente correlato alla questione esistenziale e alla preoccupazione reli-

giosa del filosofo danese di "istituire un rapporto assoluto con l'Assoluto", dato che è solo l'Assoluto ad avere proprietà di individualizzare e di determinare singolarmente l'esistente. In tal senso, oltremodo convincente risulta il modo in cui Scapolo affronta la densa ed essenziale questione kierkegaardiana del "dire l'esistenza" mediante la strategia della "verità della finzione": la pseudonimia è una *odissea personale* che permette di "prendere alle spalle" il lettore, di "ingannarlo" per "renderlo attento al cristianesimo", a una verità soggettiva ed etico-esistenziale che non si può *dire direttamente*, ma solo *mostrare* indirettamente.

Ed è proprio il carattere asintotico e approssimativo del sapere oggettivo, è proprio l'aspirazione a porre rimedio all'impossibilità di comprendere e di rappresentare, ciò che sta alla base del significato dello pseudonimo Johannes de Silentio – il "silenzioso sulla terra" (*Stille nî Landen*) a cui rinvia il pietismo – che "parla in silenzio" del silenzio di Abramo, consapevole della sua impossibilità di dire, di spiegare e *condividere* con altri la propria *esperienza limite*. In maniera inedita rispetto agli studi sull'opera, Scapolo rintraccia in tale angosciata chiusura comunicativa tutto il portato filosofico ed etico-religioso del gesto paradossale di Abramo.

È questo lo snodo teoretico dirimente dello studio di Scapolo, il concetto-chiave assunto come *Leitmotiv* e linea-guida dell'intera impostazione della sua analisi: negando la validità e l'*inter-esse* di qualsivoglia astrazione oggettivante, Kierkegaard rifiuta ogni conciliazione – la logica del correlativo

copulativo *baade-og* – ogni mediazione hegelianamente intesa (*Versöhnung*) sostituendola con il “paradosso” (*det Paradox*), “il *pathos* della vita intellettuale” che è estraneo a ogni idea di Assoluto capace di compiersi nel concetto. Sebbene non manchi di inserire la questione spinosissima e tutt’altro che risolta del paradosso nella cornice critica dei due poli entro cui si muove quasi tutta la critica che si occupa, nello specifico, di *Timore tremore* – quello che legge il paradosso come un *impasse* dell’intelletto, e quello che invece vi scorge la più alta e concreta possibilità raggiungibile dall’attività speculativa – Scapolo mira chiaramente a scardinare la tradizionale interpretazione *irrazionalista* di Kierkegaard, evidenziando come la sua passione per il paradosso e la sua avversione nei confronti della speculazione hegeliana non significhino *eo ipso* rinuncia al pensiero razionale: «Il paradosso, pietra d’inciampo di ogni riflessione, è per Kierkegaard l’unica “passione del pensiero”. Solo tramite esso l’uomo può spingere la propria riflessione al suo intrinseco limite estremo, al suo “tramonto”. [...] È attraverso il paradosso che l’uomo giunge a scoprire ciò che non si può pensare, ciò che non si può “risolvere logicamente” o “mediare razionalmente”, e che tuttavia costituisce la sostanza della sua stessa vita» (p. 67). Il paradosso è dunque – questo emerge bene dal lavoro di Scapolo – la categoria cardine sia della *dialettica dell’esistente* che si gioca nello scarto tra razionale e reale, sia della *dialettica qualitativa dell’enten-eller* fatta di decisioni e disgiunzioni, di rotture e salti, di contraddizioni e di

scandali che non si lasciano ridurre all’appropriazione meramente speculativa.

L’ermetismo della “continua lotta con la ragione” viene ulteriormente scandagliato da Scapolo nel secondo capitolo del suo lavoro (*La “bella storia di Abramo”. Analisi e commento*), dove non solo affronta il percorso retorico-letterario dell’opera, non solo si cala nel *nucleus in nuce* della lettura esegetica di *Timore e tremore*, ma ne presenta soprattutto, in maniera puntuale e dettagliata, la pregnante essenza filosofico-poetica custodita nella figura di Abramo, il Cavaliere della fede in grado di suggerire una visione precisa della “sospensione teleologica dell’etica”. Seppure sia stato assunto a paradigma dell’autentica e viscerale religiosità del credente, nel gesto esemplare di Abramo vi è però qualcosa *in più* di una semplice disposizione all’obbedienza a un comando impartito da Dio e inaccettabile agli occhi della legge umana. Con sensibilità e acutezza, Scapolo indaga proprio questo *di più*, interroga il valore primigenio dell’esperienza silenziosa di Abramo, sviscera in maniera scrupolosa e sicura la problematicità specifica di quella “felice passione” tra l’intelletto e il paradosso che Kierkegaard rintraccia nella “fede” (*Tro*).

Il gesto di Abramo è *par excellence* il gesto del *credo quia absurdum*, il gesto che non si lascia mediare perché è fede nell’assurdo. E allora Johannes de Silentio legge la storia di Abramo narrata in *Genesi* staccandosene, diventandone l’osservatore, dato che la comprensione di ciò che non si può spiegare e capire può passare solo attraverso il tentativo di *vivere* quel che non si arriva a com-

prendere. L'asse di rotazione e la chiave interpretativa principale di *Timore e tremore* consiste proprio in questa *crux mea lux* ermeneutica cui va sottoposta l'eccezionalità dell'esperienza-limite – per dirla in termini blanchotiani – della fede di un Abramo che, tuttavia, non è l'Abramo dell'Antico Testamento. Kierkegaard-de Silentio infatti propone quattro finzioni o variazioni poetiche di Abramo con l'intento di far emergere per *via negativa* la vera statura di colui che ha abbandonato i “dicibili” criteri precettistici e dogmatici in nome di una vertigine *de visu* con Dio solo, in nome di un abbandono umanamente ingiustificabile.

Scapolo scava nella eccezionalità della fede abramitica “osservata” da Johannes de Silentio e ne rintraccia la dimensione cruciale – non in linea con la fede cristiana – sia nel fatto che la semplice volontà di obbedire al comando divino non è un indicatore significativo né la garanzia sufficiente di quel Johannes de Silentio intende per fede, sia nel fatto che Abramo non spera semplicemente di ricongiungersi al figlio in un'altra vita, ma in *questa* vita. Si tratta dunque di una fede *immanente* al nostro mondo che non chiede qualcosa di lontano e di futuro, ma una “immortalità presente” nel finito e nel tempo. L'incontro di immanenza e trascendenza è il confine che distingue nettamente il Cavaliere della fede (*Troens Ridder*), qual è l'Abramo descritto in *Genesis*, e il Cavaliere della rassegnazione infinita, segno distintivo dell'eroe tragico: se nella “infinita rassegnazione” il cavaliere rinuncia e svaluta il finito per l'infinito, non riuscendo mai a riconciliarsi con il mondo, il Cavaliere della

fede – questa *declinazione-limite* del Singolo, come efficacemente lo definisce Scapolo – vive invece una gioia riferita al finito perché conserva, nella fede, l'attenzione per questa esistenza e la capacità di rimanervi. Con ragguardevole piglio critico-teoretico Scapolo condensa molto bene questo aspetto decisivo non solo puntualizzando la potenza filosofica della narrazione ipotetica che Johannes de Silentio mette in atto per tentare di approssimare col pensiero ogni valenza possibile del gesto del Cavaliere della fede, ma soprattutto chiarendo come il gesto di Abramo abbia *dissolto l'etica* (il piano del Generale) per *tener salda* la sua fede *nell'incertezza*: «[...] la figura di Abramo è l'eccezione, la declinazione-limite, la radicalizzazione più profonda della singolarità, e tale figura può essere compresa pienamente solo attraverso l'esperienza della fede, in virtù del doppio movimento (dialettico) che essa presuppone, attraverso il quale si può riconquistare il finito passando per l'infinito» (p. 136). Questa fede nell'assurdo, che dunque consiste nel credere fermamente *l'impossibile possibilità*, si realizza mediante il gesto di Abramo che sfida, nell'istante-esemplare in cui egli *ignora* se egli stesso sia un Cavaliere della Fede o un omicida, il *mysterium tremendum* e l'abisso dello scarto incolmabile stabilito nella relazione assoluta con l'Assoluto.

L'itinerario di lettura e analisi di *Timore e tremore* indicato da Scapolo è senza dubbio considerevole, anche in ragione del fatto che, senza alcuna pretesa di esaustività, la studiosa riesce a sollecitare interrogativi stringenti rispetto all'opera kierkegaardiana

e rispetto a concetti-chiave problematici e oscuri della sua *Weltanschauung* filosofico-religiosa: cosa significa veramente “avere fede in forza dell’assurdo”? Basta che qualcosa sia assurdo perché l’uomo venga indotto a credere in esso? E inoltre – per riprendere una obiezione mossa a Kierkegaard dal teologo islandese Magnus Eiriksson, autore (sotto lo pseudonimo di Theophilus Nicolaus) di un saggio del 1850 dal titolo *La fede è un paradosso ed è “in forza dell’assurdo”*? – la fede di Abramo è veramente una fede “in forza dell’assurdo”? Non è forse vero che Abramo confida in Dio, nonostante la paradossalità e l’assurdità di quanto gli è stato chiesto, perché confida nelle promesse fattegli da Dio? La fede che si fonda sulla certezza che “per Dio non c’è niente di impossibile”, è veramente una fede “in forza dell’assurdo”? E questa fede che confida nelle promesse di Dio, sebbene non riesca a comprendere i Suoi piani, non ha forse un fondo di razionalità?

Con *Timore e tremore*, infine, Kierkegaard-Johannes de Silentio provoca una riflessione radicale e per certi aspetti controfattuale rispetto alla religiosità del nostro tempo che, dominato dalla scienza e dalla tecnica, è sempre più diffidente rispetto a un intervento miracolistico di Dio, e sempre più bisognoso di avvicinarsi all’immagine di un Dio comprensivo e che scaldi il cuore dell’uomo. Non è forse un caso che lo stesso Kierkegaard abbia preso chiaramente le distanze, nella seconda parte del suo itinerario di pensiero, dall’esaltazione della figura di Abramo fatta da Johannes de Silentio, fino a dare un giudizio più negativo sull’episodio

biblico del sacrificio di Isacco. Non solo: in *Timore e tremore* Kierkegaard ci ricorda che la religione, nella misura in cui ci induce a vivere l’oggi come dono, costituisce il migliore antidoto contro le angosce e le preoccupazioni per il domani. Sembra dunque questo il Kierkegaard che oggi vorremmo sentire più vicino, il Kierkegaard che non è affatto ostile all’umano *tout court*, che non presenta l’uomo di fede come colui che è necessariamente condannato a soffrire, ma come colui che possa anche godere delle piccole gioie e dei semplici piaceri dell’esistenza.

Alessandra Granito

Abstracts and Indexing

Author: Tommaso Reggiani

Affiliation: Universität zu Köln

Contact: tommaso.reggiani@gmail.com

Title: *Una breve nota sull'etica comportamentale*

Abstract: The traditional method of research and teaching in philosophical field adopts a purely Normative Approach . It focuses on the question on how a person should be behaving when exposed in front of ethical dilemmas . In this note we describe the principles of ethical traditional and then proceed to provide examples of how the positive approach of behavioral ethics may enrich profitably classic perspectives . In this context the ethical behavior is outlined as the study of which behaviors systematically and predictably diverge with respect to the intuition. Focusing on a positive - descriptive dimension, rather than a purely normative, the behavioral analysis of the ethical phenomena is especially interesting to investigate causes and determinants inducing motivated people to break the moral and ethical standards . The considerations and recommendations deriving from this approach are being particularly interesting and vastly applicable to concrete cases of business ethics

Keywords: Behavioural Ethics, Normative and positive approach

Location: *Philosophical Readings*, VI.3 (2014), pp. 5-8

Author: Matteo Motterlini – Carlo Canepa

Affiliation: Cresa, Università San Raffaele

Contacts: ca.canepa@gmail.com;
matteo.motterlini@univr.it

Title: *La rivoluzione "gentile" delle politiche basate sull'evidenza. Considerazioni epistemologiche*

Abstract: Is it possible to design public policies

which are thought and implemented upon citizen's preferences ? Behavioral approaches to policy making aim to create environments which improve decisions , based on the complexity of social and cognitive factors that affect the decision process . Their implementation is guided by the evidence of the results obtained once policies are tested . In this way the institutions and policy makers have one tool to promote a civic behavior for individual citizens and the community expanding freedom of choice and simplifying the regulatory policies . This article examines the epistemology of planning a correct choice architecture.

Keywords: Evidence Based Policy; nudge; choice architecture

Location: *Philosophical Readings*, VI.3 (2014), pp. 9-23

Author: Alessandra Smerilli

Affiliation: Università Pontificia Salesiana

Contact: alessandro.smerilli@gmail.com

Title: *Theories of Team Reasoning*

Abstract: We-thinking theories allow groups to deliberate as agents. They have been introduced into the economic domain for both theoretical and empirical reasons The idea of we-thinking, or we-reasoning, is increasingly drawing the attention of more and more economists. Within this body of literature, Robert Sugden and Michael Bacharach have developed analytical frameworks from an economic point of view. The main claim of scholars that analyze we-thinking is that it can be endorsed by people when they face a decision problem. In fact, experimental evidence shows that, especially in some kind of games, such as coordination games, people do endorse we-thinking. However, the way in which we-thinking arises and how it brings people to behave in a particular way in games is a matter that requires further investigation.

Keywords: team reasoning, we thinking

Location: *Philosophical Readings*, VI.3 (2014), pp. 24-34

Author: Daniela Greco

Affiliation: Università Bocconi

Contact: daniela.grieco@unibocconi.it

Title: *On the Demand and Effectiveness of Reward in Voluntary Contribution Mechanisms. A Note*

Abstract: This note aims at discussing the reasons why individuals facing situations of anonymous interaction with strangers rarely decide to remunerate pro-social behaviours. Moreover, economic experiments show that, when implemented, rewards are quite ineffective in promoting cooperation. Two key explanations are discussed: rewarding does not enable to yell out the negative emotions that instead can be expressed by assigning a sanction. On the other hand, rewards are received by subjects holding an intrinsic motivation to contribute and this could undermine the effectiveness of an extrinsic motivation like reward.

Keywords: Public Good Games; Reward; Punishment; Experimental Economics; Cooperation

Location: *Philosophical Readings*, VI.3 (2014), pp. 35-45