

Regole. Un saggio di psicodeontica

Marco Brigaglia*

Università degli Studi di Palermo

Abstract: Rules. An Essay in Psychodeontics

In this article, I first introduce a naturalistic method of inquiry in meta-normative issues, that, following a suggestion by B. Celano, I call “psychodeontics”. I then provide an example of how the method works, focusing on three aspects of the phenomenon of rules and rule-guided behaviour: the fact that rules are reasons for action, the relation between rules and habits, and the “entrenchment” of rules.

Keywords: Naturalization of normativity, Rules, Psychology of rules, Rule-Guided Behaviour.

Sommario: 1. Premessa – 2. Naturalizzazione meccanicista – 3. Struttura della spiegazione meccanicista – 4. Il ruolo della filosofia e la psicodeontica – 5. Regole: alcuni aspetti del fenomeno – 6. Controllo cognitivo e automaticità – 7. Struttura del controllo cognitivo – 8. Controllo cognitivo e ragioni – 9. Regole e abitudini nel cervello – 10. Regole trincerate e meccanismi di *gating*.

1. Premessa

Il termine “psicodeontica” è stato introdotto qualche anno fa da Bruno Celano¹ per indicare tre cose interconnesse: un campo di studi, l’insieme di assunti ontologici e metodologici ad esso inerenti, e, infine, un programma di ricerca fondato su quegli studi e quegli assunti.

Il campo di studi è quello delle indagini empiriche, condotte nell’ambito delle scienze cognitive (in primo luogo, psicologia sperimentale e neuroscienze), sui fenomeni normativi in genere, e in particolare su regole e comportamenti guidati da regole.

Questo campo di studi, argomentava Celano, è, consapevolmente o inconsapevolmente, guidato da un insieme di assunti di stampo naturalista o

* La ricerca che ha condotto a questo articolo è stata co-finanziata dal “Premio giovani” del Dipartimento di Giurisprudenza dell’Università di Palermo.

¹ B. Celano, “Ragionamento giuridico, particolarismo. In difesa di un approccio psicologico”, in *Rivista di Filosofia del diritto*, 2017, n. 2, pp. 315-344.

“psicologista” (Celano ha scelto di adottare questa etichetta *démodé*, come provocazione contro la sua resistente, ma a suo parere antiquata, connotazione negativa). Gli assunti sono, in estrema sintesi, i seguenti: le inferenze pratiche e i criteri di correttezza che le guidano non sono altro che processi ed entità psicologiche; un’adeguata comprensione della loro struttura e funzionamento non può essere ottenuta *solo* attraverso metodi filosofici tradizionali, ma richiede *anche* indagini empiriche.

Il programma di ricerca è l’esplorazione di problemi di teoria del diritto e del ragionamento pratico alla luce di quel campo di studi e di quegli assunti – una forma di naturalizzazione. È un programma che progettavamo di sviluppare insieme, ma insieme abbiamo avuto il tempo di fare appena un paio di passi².

“Psicodeontica” era una etichetta di comodo, che Bruno non prendeva del tutto sul serio – mentre prendeva molto sul serio il campo di studi, gli assunti e il programma di ricerca che essa designa. Qui, per ricordarlo, vorrei invece prendere sul serio anche l’etichetta, e rilanciarla come nome adatto a indicare un modo specifico di procedere – un vero e proprio metodo – nella strada verso la naturalizzazione della normatività.

Nella prima parte del saggio (§§ 2-4), offrirò un inquadramento più preciso del metodo psicodeontico, collocandolo all’interno di una particolare versione del programma di naturalizzazione, che chiamerò “naturalizzazione meccanista”. Nella seconda parte del saggio (§§ 5-10), darò un esempio di applicazione del metodo psicodeontico centrata proprio sulla nozione di regole.

Se Bruno fosse stato qui, avrei discusso a lungo con lui le idee di questo saggio. In qualche modo l’ho fatto lo stesso, in un continuo dialogo immaginario. E quindi, ancora una volta, lo ringrazio per esserci stato sempre.

2. Naturalizzazione meccanista

Negli ultimi decenni, è emersa in filosofia della scienza una corrente nota come “neo-meccanicismo”, o filosofia neo-meccanista³. Il neo-meccanicismo è stato inizialmente proposto come razionalizzazione del modello di spiegazione – alternativo al modello nomologico⁴ – adottato nel campo delle scienze biologiche

² M. Brigaglia, B. Celano, “Reasons, Rules, Exceptions: Towards a Psychological Account”, in *Analisi e diritto*, 2018, pp. 131-144; M. Brigaglia, B. Celano, “Constitutive Rules: The Symbolization Account”, in *Ratio Juris*, 34 (2021), n. 3, pp. 241-262.

³ Cfr. S. Glennan, *The New Mechanical Philosophy*, Oxford University Press, Oxford, 2017; C.F. Craver, G. Tabery, “Mechanisms in Science”, in *Stanford Encyclopedia of Philosophy*, 2019, <https://plato.stanford.edu/archives/sum2019/entries/science-mechanisms/>

⁴ W. Bechtel, A. Abrahamsen, “Explanation: A Mechanistic Alternative”, in *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36 (2005), pp. 421-441.

e, soprattutto, delle neuroscienze cognitive, dove ha ormai acquisito una posizione dominante⁵.

In via generale, un “meccanismo” è una struttura costituita da un insieme di componenti fisiche, organizzate in modo tale che le loro operazioni e interazioni causali producano un certo “fenomeno”, un pattern o evento osservabile. Un meccanismo è stratificato quando i suoi componenti sono, a loro volta, meccanismi.

La nozione cruciale per le scienze cognitive è quella di meccanismo “mentale”⁶. I meccanismi “mentali” sono meccanismi che regolano dinamicamente l’interazione (causale) tra un organismo vivente e l’ambiente sociale e naturale in cui è immerso – o interazioni tra parti dell’organismo, funzionalmente connesse all’interazione con l’ambiente –, attraverso le operazioni (causali) di una specifica struttura organizzata parte dell’organismo stesso, il sistema nervoso centrale⁷. I meccanismi mentali sono stratificati, scomponibili in meccanismi – organizzazioni – di livelli via via inferiori: circuiti di neuroni, singole celle neuronali, molecole, atomi, ecc. I meccanismi mentali, infine, sono strutture “fisiche”: procedendo di livello in livello nella scomposizione meccanicista, essi risultano infatti, in ultima analisi, interamente composti dalle stesse entità di base, operanti secondo le stesse leggi naturali, che compongono le cose prive di mente, le entità “fisiche” nel senso più familiare del termine⁸.

La filosofia meccanicista supporta una forma di naturalizzazione molto diffusa nelle neuroscienze cognitive, che chiamerò “naturalizzazione meccanicista”. La naturalizzazione meccanicista di un fenomeno mentale target comprende due passaggi. Il primo passaggio è una congettura di stampo fisicalista: che quel fenomeno non sia niente altro che un pattern o un evento di interazione fra l’organismo e l’ambiente (due entità fisiche), o tra parti dell’organismo, orchestrato da un certo insieme di meccanismi mentali (altre entità fisiche); non sia, cioè, niente altro che organizzazione, stratificata su vari livelli, di entità fisiche. Il secondo passaggio è il supporto di questa congettura attraverso lo sviluppo di una spiegazione meccanicista del fenomeno mentale target. La spiegazione meccanicista parte da una descrizione iniziale del fenomeno, e mira all’elaborazione di un modello scientificamente plausibile, e specificato a livello neurale, dei meccanismi mentali soggiacenti.

L’elaborazione di un modello del genere è un compito che spetta alle neuroscienze cognitive. Ma la spiegazione meccanicista nel suo complesso è

⁵ V. C. Craver, *Explaining the Brain. Mechanisms and the Mosaic Unity of Neuroscience*, Oxford University Press, Oxford, 2007; W. Bechtel, *Mental Mechanisms. Philosophical Perspectives on Cognitive Neuroscience*, Routledge, New York, 2008.

⁶ V. soprattutto W. Bechtel, *op. cit.*, su cui mi baserò nelle pagine seguenti.

⁷ Questa è una nozione ristretta di meccanismo “mentale”. Per i fan della realizzabilità multipla e della mente estesa, resta aperta la possibilità di ampliare la nozione in modo da includere anche meccanismi che svolgono funzioni analoghe in organismi diversi, come per esempio organismi artificiali, o organismi viventi integrati da supporti artificiali.

⁸ Per questa nozione di “fisico” v. J. Prinz, *The Conscious Brain. How Attention Engenders Experience*, Oxford University Press, Oxford, 2012, p. 11.

un'impresa più ampia, di natura multidisciplinare. Entro di essa, le neuroscienze cognitive costituiscono non tanto l'apice, quanto piuttosto il punto di snodo tra ambiti di conoscenza e ricerca sulla mente, e sulla costituzione del suo supporto biologico, molto diversi per tradizioni e metodi. Da un lato, vi sono tipiche scienze naturali che operano a livelli di organizzazione e descrizione più bassi (il livello della "materia" nervosa), seguendo la decomposizione dei meccanismi mentali giù verso le loro componenti fisiche di base (neurofisiologia, neuroanatomia, biologia cellulare e molecolare, biochimica). D'altro lato, vi sono conoscenze e discipline che operano a livelli di descrizione più alta (il livello dello "spirito"). Fra esse vi è anzitutto la psicologia, sia la psicologia scientifica (una scienza la cui collocazione tra scienze naturali, sociali e umane è stata ambigua e controversa), sia la psicologia di senso comune (*folk psychology*), il vasto insieme di modelli pre-teorici della mente e delle sue attività condivisi in un certo contesto culturale. Ma vi è anche la riflessione filosofica sulla mente. Per introdurre il suo ruolo, è necessario osservare più da vicino la struttura della spiegazione meccanicista.

3. Struttura della spiegazione meccanicista

In un modello idealizzato e drasticamente semplificato, la spiegazione meccanicista procede così.

Il punto di partenza è la descrizione del fenomeno target, che prende inevitabilmente le mosse dalla psicologia di senso comune. Un ottimo esempio è la spiegazione meccanicista della memoria⁹. Nell'accostarsi alla memoria come oggetto di indagine, lo scienziato muove anzitutto da una prima concettualizzazione del fenomeno sulla base della psicologia di senso comune, che comprende, fra l'altro: la discriminazione stessa della memoria come processo mentale specifico, distinto da altri processi mentali, come la percezione o la decisione; l'individuazione di una serie di esperienze prototipiche di fenomeni mnemonici (dimenticare o ricordare, tenere pronto in mente o memorizzare, sapere di sapere ma non riuscire a verbalizzare, ecc.); un vasto insieme di tacite assunzioni teoriche riguardo ai fenomeni mnemonici (ad esempio, il fatto che la memoria tende a degradarsi nel tempo, o il fatto che abbia una capacità limitata); una serie di metafore attraverso le quali è concettualizzata la memoria (la memoria come "deposito" in cui vengono "immagazzinate" le informazioni, il passato che "torna" alla mente, qualcosa che è "sulla punta della lingua", ecc.). Tutto questo insieme di conoscenze costituisce già una prima descrizione del fenomeno.

La psicologia scientifica interviene rimodellando la pre-comprensione di senso comune del fenomeno, sulla base e in vista di indagini empiriche. La rappresentazione di senso comune della memoria viene anzitutto raffinata introducendo nuove distinzioni: memoria a lungo e breve termine e memoria di

⁹ L'esempio è quello proposto da W. Bechtel, *op. cit.*, cap. 2.

lavoro (uno “spazio mentale”, di capacità limitata, dove l’informazione è mantenuta in uno stato attivo per guidare contestualmente l’azione e il pensiero); all’interno della memoria a lungo termine, memoria implicita e memoria dichiarativa; all’interno di quest’ultima, memoria semantica ed episodica, ecc. L’elaborazione di queste distinzioni è guidata dall’indagine empirica, e orientata ad essa. Le distinzioni sono infatti suggerite da rilevazioni empiriche (ad esempio, l’osservazione che certe lesioni cerebrali si associano a deficit di quella che poi sarà chiamata memoria dichiarativa ma non di quella che poi sarà chiamata memoria implicita), e sono inoltre “operazionalizzate”, costruite in modo tale da poter essere indagate sperimentalmente. Infine, vengono formulate ipotesi riguardo ai meccanismi responsabili della memoria, nelle sue diverse forme. Queste ipotesi sono fraseggiate in termini psico-funzionali, e cioè scomponendo le operazioni inerenti ai processi mnemonici in operazioni più semplici (un meccanismo addetto all’immagazzinamento dell’informazione, un altro addetto al suo recupero, ecc.). Queste spiegazioni, però, hanno carattere “omuncolare”. Non viene cioè specificato come le operazioni individuate possano essere realizzate da strutture neurali sub-personali, non dotate di intelligenza. Semplicemente, si postula che lo siano – come se vi fosse un omuncolo deputato alla loro esecuzione.

Le neuroscienze cognitive utilizzano, come punto di partenza, i modelli elaborati dagli psicologi, cercando di sviluppare una spiegazione capace di eliminare gli “omuncoli” sostituendoli con processi neurali specifici, capaci di svolgere le funzioni appropriate. Ma il confronto con il vincolo biologico è tutt’altro che passivo: man mano che la conoscenza delle strutture e funzioni neurali aumenta, i modelli esplicativi di partenza sono rivisti anche pesantemente. Un ottimo esempio è quello della ricerca sulla memoria di lavoro, che sfioreremo più avanti. Nei modelli originari, l’accesso e il mantenimento dell’informazione nella memoria di lavoro erano coordinati da un “esecutivo centrale” (un omuncolo). Nei modelli attuali – fondati sull’architettura del cervello e sull’individuazione dei processi cerebrali attivi durante l’esecuzione di un compito, oltre che su simulazioni con network neurali artificiali –, l’idea dell’esecutivo centrale è stata abbandonata, in favore di network neurali complessi in cui l’interazione causale delle parti, in assenza di qualsiasi coordinamento centrale, è in grado di svolgere le funzioni appropriate.

Un’ultima, importante questione riguarda il destino della descrizione iniziale di senso comune del fenomeno. Essa non costituisce, durante il processo di spiegazione meccanicista, un dato fisso, che può solo essere precisato e arricchito, ma non messo in questione. Al contrario, oltre che poco accurata essa può anche rivelarsi parzialmente inadeguata, falsa. Ad esempio, la psicologia di senso comune rappresenta la memoria come registrazione veridica di eventi. Ma molta ricerca psicologica e neuroscientifica mette in dubbio questa concezione, suggerendo invece che la memoria sia un processo radicalmente costruttivo: non recuperiamo ricordi già pronti, ma li ricostruiamo volta per volta. Il fenomeno target, a questo punto, appare diverso rispetto a come era stato inizialmente concepito – la spiegazione meccanicista non ha semplicemente individuato i meccanismi

responsabili di un fenomeno *dato*, ma ha “riconfigurato” il fenomeno stesso. Questa riconfigurazione può, in certi casi, generare una discrasia tra l’immagine scientifica della mente e il senso comune. Oppure, la riconfigurazione suggerita dall’indagine scientifica può retroagire sul senso comune, che “co-evolve” insieme ad essa¹⁰ – ciò accadrebbe, ad esempio, se cominciasimo comunemente a rappresentarci la memoria come ricostruttiva.

Sebbene molto semplificata rispetto alle reali dinamiche esplicative, spero che questa breve ricostruzione sia riuscita a trasmettere il punto centrale: la *circolarità* fra i diversi livelli della spiegazione meccanicista. È in questa circolarità che si inserisce l’indagine filosofica.

4. Il ruolo della filosofia e la psiconeontica

A seguito del dibattito sullo psicologismo infuriato nelle università tedesche fra il 1880 e il 1920¹¹, per gran parte del Novecento le indagini sulla mente di ambito filosofico sono rimaste programmaticamente separate dalla psicologia empirica, e condotte con metodi diversi. I filosofi si sono concentrati sull’osservazione introspettiva e la concettualizzazione dell’esperienza soggettiva, e sulla delucidazione, esplicitazione o ricostruzione razionale dei concetti della psicologia di senso comune, rifiutandosi, per lo più, di integrare questi metodi – “tradizionali” perché percepiti come continui rispetto alla tradizione filosofica – con i “nuovi” metodi empirici messi a punto da altre scienze naturali e adottati dagli psicologi – esperimenti in condizioni controllate e analisi statistiche.

Non è importante, in questa sede, soffermarsi sulle giustificazioni ontologiche, epistemologiche e meta-filosofiche di queste (implicite o esplicite) scelte metodologiche. Il punto centrale, qui, è un altro. Le elaborazioni “tradizionali” – o gran parte di esse – costituiscono un *prolungamento riflessivo* (e, a volte, una problematizzazione) della psicologia di senso comune, che può innestarsi perfettamente nella spiegazione meccanicista. Esse, infatti, forniscono descrizioni iniziali dei fenomeni da indagare, e ipotesi (implicite o esplicite) sui meccanismi sottostanti, molto più fini e articolate rispetto a quelle sottese al senso comune, e molto più capaci di ispirare e orientare la ricerca delle scienze cognitive. Un esempio celebre sono le riflessioni di Wittgenstein sulle “somialtanze di famiglia”, che hanno ispirato le ricerche psicologiche sulla struttura prototipica dei concetti.

Uno dei compiti del filosofo naturalista diventa a questo punto ricordare l’immenso bacino di conoscenza della filosofia “tradizionale” con l’indagine psicologica e neuroscientifica, in vista della spiegazione meccanicista. Ciò significa,

¹⁰ P.S. Churchland, *Brain-wise. Studies in Neurophilosophy*, The MIT Press, Cambridge (MA), 2002.

¹¹ Per una ricostruzione v. M. Kusch, *Psychologism. The Sociology of Philosophical Knowledge*, Routledge, London, 1995.

anzitutto, fare emergere parziali corrispondenze tra le descrizioni dei fenomeni e il relativo linguaggio concettuale messe a punto dai filosofi e quelle in uso in ambito scientifico (ne vedremo più avanti un esempio con la nozione di “controllo cognitivo” e la nozione di “ragioni”). Significa tradurre teorie filosofiche riguardo ad un certo fenomeno in ipotesi, suscettibili di conferma empirica, sulle strutture dei processi mentali soggiacenti. Significa fare emergere parziali corrispondenze tra questi modelli esplicativi di alto livello, e le analisi meccaniciste sviluppate a livello psicologico e neuroscientifico. Significa, ancora, “riconfigurare” la descrizione filosofica dei fenomeni target in modo da armonizzarla con i modelli meccanicisti delle scienze cognitive.

La psicodeontica, per come la intendo, è questo lavoro di raccordo, riferito al fenomeno della normatività. Si tratta, chiaramente, di un campo di studi vastissimo, già ampiamente in corso, dalla psicologia morale agli studi psicologici sul ragionamento. Ma il lavoro da svolgere è ancora tantissimo. Uno dei campi meno esplorati riguarda proprio alcuni aspetti della nozione (o meglio, di *una* nozione) di regole e di comportamento guidato da regole messa a punto da teorici del diritto, che sarà discussa nei prossimi paragrafi.

Vi sarebbero molte cose da aggiungere riguardo al programma di naturalizzazione meccanicista, per rispondere alle tipiche obiezioni di cui questa e altre forme di naturalismo sono oggetto. Bisognerebbe, in particolare, insistere su fatto che la naturalizzazione meccanicista muove da una *congettura* fiscalista, non necessariamente generale, ma limitata ad *un fenomeno target*. Bisognerebbe mostrare come la naturalizzazione meccanicista sia riduzionista solo in un senso innocuo rispetto al pluralismo metodologico, perché non elimina affatto, ma anzi enfatizza, la molteplicità di livelli di descrizione dei fenomeni. Bisognerebbe, ancora, spiegare come la naturalizzazione meccanicista non neghi affatto la rilevanza della distinzione fra prospettiva di prima e di terza persona. Bisognerebbe, infine, enfatizzare come il programma psicodeontico riguardi un riorientamento della teoria del diritto *complessivamente considerata*, e non del lavoro di ciascun ricercatore – alcuni continueranno legittimamente a usare metodi tradizionali, mentre altri si occuperanno del raccordo psicodeontico. Purtroppo, però, non ho spazio sufficiente per tutto questo. Sarà per un'altra volta.

5. Regole: alcuni aspetti del fenomeno

Il fenomeno di cui mi occuperò qui è quello delle regole intese in senso normativo, come guida del comportamento. La nozione normativa di regola è senza dubbio parte della psicologia di senso comune (fa parte del senso comune l'idea che le regole possano essere apprese, memorizzate, richiamate alla mente, e usate per decidere come comportarsi). Più che da definizioni esplicite, la nozione è catturata da innumerevoli *platitudes* e da convergenze sistematiche nella sua applicazione, governate, verosimilmente da una rete largamente condivisa di prototipi e assunzioni teoriche tacite.

Le regole e il loro ruolo nella guida del comportamento sono state anche un tradizionale tema di indagine per i filosofi, divenuto, nel secolo scorso, oggetto di ricerca particolarmente intensa e sofisticata, soprattutto nell'ambito della teoria del diritto e del ragionamento pratico. Ne sono risultati innumerevoli tentativi di esplicitazione, delucidazione, precisazione, ricostruzione, elaborazione, e a volte messa in questione del fenomeno delle regole per come concepito dal senso comune. Visto dal punto di vista della spiegazione meccanicista, si tratta di un ricchissimo bacino da cui trarre descrizioni di grana fine degli aspetti più vari del fenomeno, capaci di orientare l'indagine psicologica e neuro-scientifica in modo molto più specifico del mero senso comune, e spesso carichi, più o meno esplicitamente, di ipotesi interessanti sui meccanismi soggiacenti.

Qui, mi soffermerò su tre punti divenuti ormai parte integrante della rappresentazione standard delle regole nella teoria del diritto contemporanea. Mi rifarò principalmente alle elaborazioni di H.L.A Hart, J. Raz e F. Schauer. Il mio intento però, è il caso di rimarcarlo, non è esegetico. Ciò che mi preme fare è, semplicemente, usare alcune idee e modelli tratti dal, o ispirati al, lavoro di questi autori per descrivere i tre aspetti in esame in un modo che guarda già alle spiegazioni meccaniciste delle scienze cognitive, con l'intento di mettere in luce punti sia di contatto che di divergenza tra elaborazioni filosofiche e indagini scientifiche.

(1) *Regole come ragioni d'azione*. Il primo aspetto riguarda la natura di "ragione d'azione" delle regole. Come è noto, è Hart ad averlo (non certo scoperto ma) portato in primo piano¹². Il senso normativo (non unico ma) più interessante in cui il termine "regola" è comunemente inteso, spiega Hart, è proprio questo: un "modello di condotta" – e cioè la rappresentazione astratta di un pattern di comportamento possibile, una condotta-tipo in una situazione-tipo – che ha il ruolo di *ragione d'azione*.

Che tipo di entità siano le ragioni d'azione è controverso. Ma vi è largo consenso sui due versanti principali, strettamente connessi, del ruolo da esse svolto: la *guida normativa* dell'azione, e la sua *giustificazione*. Qui, mi concentrerò sul primo.

Dell'idea di guida normativa dell'azione, Hart mette a fuoco un aspetto centrale, il *controllo intenzionale dell'azione sulla base di standard coscientemente rappresentati* (e socialmente condivisi – di questo, però, non mi occuperò). "Controllare" intenzionalmente l'azione significa non lasciare che sia determinata da meri impulsi, abitudini, o spinte immediate e prepotenti, ma sottometerla invece ad un processo di riflessione, che può suggerire di non compiere l'azione verso la quale incliniamo in modo più immediato, e di compiere un'azione diversa. Questo processo di riflessione può essere risolto in modo intuitivo – un corso d'azione ci appare preferibile, e continua ad apparirci tale dietro riflessione, ma non sappiamo in virtù di cosa. Oppure, può essere risolto in modo raziocinativo: scegliamo un

¹² H.L.A. Hart, *The Concept of Law*, Clarendon Press, Oxford, 1961.

corso d'azione *proprio in virtù* della sua appropriata connessione con un certo standard, di cui siamo consapevoli. Un certo standard P conta come ragione d'azione per un soggetto X, se X sceglie o tende a scegliere A *proprio in virtù* della sua appropriata connessione con P, di cui X è consapevole.

La via maestra per i tentativi di naturalizzazione delle ragioni passa attraverso due mosse. Gli “standard” sono intesi come entità psicologiche, rappresentazioni mentali – in particolare, nel caso delle ragioni “operative”¹³, sono rappresentazioni motivanti, “pro-attitudini” (internalismo). Il processo che conduce dalla rappresentazione cosciente dello standard alla scelta è un processo causale (causalismo). Il compito della spiegazione meccanicista è, in questa prospettiva, specificare di che tipo di rappresentazioni mentali si tratti e attraverso quali meccanismi esse influenzino causalmente la scelta.

Vi sono, però, varie difficoltà. La prima difficoltà è che non ogni rappresentazione mentale che influenza causalmente la scelta conta per noi come una ragione – la nozione di ragione sembra essere molto specifica, e la spiegazione meccanicista deve render conto di questa specificità.

La seconda difficoltà è che il concetto di ragione ha una componente ineludibilmente normativa. P è, in senso pieno, una ragione per X, se X tende a scegliere A riconoscendo, o essendo disposto a riconoscere, che A è connessa con P *in modo appropriato*, e che è *corretto* accettare P come ragione. P è, in breve, una ragione per X, se X la riconosce, o è disposto a riconoscerla, come una *buona ragione*, e non invece a squalificarla come una mera apparenza di ragione – una mera *causa*. È possibile dar conto in termini meccanicisti di questa componente normativa?

In questa sede, mi limiterò a dire qualcosa sui due problemi su indicati. Ma essi non esauriscono, sia chiaro, le difficoltà. L'internalismo e il causalismo riguardo alle ragioni possono essere criticati anche sotto altri punti di vista. Oltre alla dimensione soggettiva qui considerata (l'essere una ragione per un certo soggetto), le ragioni hanno anche, secondo alcuni, una dimensione oggettiva (essere una ragione, *full stop*). Se si volesse perseguire una completa naturalizzazione delle ragioni, bisognerebbe affrontare tutti questi problemi, e altri ancora.

(2) *Regole e abitudini*. *The Concept of Law* è imperniato, com'è noto, sulla distinzione fra regole e abitudini¹⁴. Essere guidati da regole e agire per abitudine sono, incontestabilmente, due diverse modalità di comportamento. Ma la relazione fra di esse – come Hart stesso implicitamente riconosce¹⁵ – può essere molto più complessa di una semplice opposizione. Il funzionamento delle regole può assumere, sotto molti aspetti, tratti abituali. Si pensi, ad esempio, al recupero di una regola dalla memoria nella situazione appropriata: con la pratica, il recupero diventa

¹³ Per la nozione di ragione “operativa” v. J. Raz, *Practical Reasons and Norms*, Oxford University Press, Oxford, 1975, p. 33 s.

¹⁴ *Ivi*, pp. 55-57.

¹⁵ *Ivi*, pp. 140 s.

sempre più facile e veloce, finché la regola non viene in mente da sola – una sorta di abitudine o abilità mentale. Si pensi, ancora, a strutture di regole, tali che il pensiero di R richiama immediatamente e regolarmente in mente anche R1 ed R2 – un'altra abitudine o abilità mentale. Ma si pensi, soprattutto, alla graduale transizione, con la pratica, dal comportamento guidato da regole al comportamento abituale – la regola passa sullo sfondo, lasciando operare una abitudine. Il fenomeno della continuità tra regole e abitudini è stato messo a fuoco da molti filosofi con semplici metodi introspettivi¹⁶. Ma l'analisi meccanicista condotta dalle scienze cognitive ne permette un inquadramento molto più preciso e realistico, e, a volte, ne può suggerire la riconfigurazione.

(3) *Regole “trincerate”*. Le regole, pur essendo usualmente giustificate in virtù della loro connessione con ragioni ulteriori (ad esempio, scopi rispetto ai quali l'osservanza della regola è strumentale), sono in grado di funzionare come ragioni in modo autonomo, per forza propria, senza bisogno di esplicitare la loro giustificazione. Questa è una caratteristica propria di tutte le regole, anche le più deboli, le cosiddette *rules of thumb*. Alcune regole, però, hanno una forza supplementare, che i filosofi hanno tentato di catturare attraverso diverse nozioni. Le più fortunate sono le *protected reasons for action* di J. Raz¹⁷, e le *entrenched prescriptive generalizations* di F. Schauer¹⁸. Le proposte di Raz e Schauer hanno dato vita ad un ricchissimo dibattito filosofico, ma, per quanto ne so, non hanno ancora avuto eco in ambito psicologico e neuroscientifico. Eppure, il fenomeno che esse tentano di inquadrare è estremamente importante, e ha dei chiari agganci con i cosiddetti meccanismi di *gating* (si noti il ricorso alla stessa metafora!) attraverso i quali le neuroscienze cognitive spiegano il *trade-off* fra flessibilità e stabilità assicurato dalle regole. Vedremo più avanti di cosa si tratta. Per adesso, è sufficiente offrire una prima descrizione del fenomeno messo a fuoco da Raz e Schauer – in termini parzialmente diversi da quelli da loro impiegati.

La guida esercitata dalla regola, e la sua capacità di influenzare il comportamento, possono essere minacciate non soltanto da impulsi prepotenti o ragioni confliggenti, ma anche da processi di ragionamento che mettono in questione la regola. Ci si può chiedere, ad esempio, se sia corretto applicare la regola senza sottoporre a scrutinio il suo “merito”, la bontà o appropriatezza del suo contenuto; oppure, ci si può chiedere se l'azione indicata dalla regola sia, nel caso particolare o in una certa classe di casi, la migliore alla luce di certi scopi o valori.

Mettere in questione la regola non necessariamente incide sulla guida che essa esercita sull'azione. Almeno in teoria, è possibile “sdoppiare” i propri processi di

¹⁶ Un esempio per tutti sono le acute considerazioni in J. Searle, *The Construction of Social Reality*, Free Press, New York, 1995, cap. 6.

¹⁷ J. Raz, *The Authority of Law. Essays on Law and Morality*, Clarendon Press, Oxford, 1979, p. 17 s.

¹⁸ F. Schauer, *Playing by the Rules. A Philosophical Examination of Rule-Based Decision-Making*, Clarendon Press, Oxford, 1991.

riflessione. Da un lato, la regola viene messa in questione. D'altro lato, si continua ad essere guidati dalla regola, schermando il proprio processo decisionale dall'interferenza delle considerazioni contro la regola. Si pensi, ad esempio, ad un processo decisionale di questo tipo: “Mi sembra davvero una regola stupida, e non ha molto senso seguirla. Ma è pur sempre una regola, e devo seguirla anche se mi sembra stupida”. Dirò che, in casi come questo, la regola è messa in questione “off-line”: le considerazioni contro la regola sono intrattenute in un modo che non ha impatto sul processo decisionale effettivo¹⁹.

Può invece accadere che, nel mettere in questione la regola, si sospenda anche il suo ruolo di guida e dunque la sua capacità di influenzare il comportamento, e si avvii invece un processo decisionale guidato da altre ragioni – ad esempio, direttamente guidato dagli scopi a cui la regola era associata come mezzo. In casi del genere, dirò che la regola non è semplicemente messa in questione “off-line”, ma è *riconsiderata*. La riconsiderazione della regola può condurre a confermare il verdetto che essa esprime, o a respingerlo, in tutto o in parte. Se, a seguito di riconsiderazione, una regola R rimane in piedi, ma viene fatta una eccezione in un caso particolare, R è *disapplicata*. Quando invece, a seguito di riconsiderazione, R è riconnessa ad una nuova regola R1 che sottrae una classe di casi alla conseguenza prevista da R, R è *specificata*.

Dico che una regola R è *trincerata*, per un soggetto X, quando in X sussistono una serie di dispositivi cognitivi che proteggono R dall'essere messa in questione “off-line” (sempre che ciò sia possibile) o riconsiderata²⁰. I dispositivi cognitivi presi usualmente in considerazione nella discussione filosofica sul fenomeno del trinceramento sono di natura raziocinativa. Consistono, essenzialmente, in ragioni: regole di livello superiore che proibiscono di mettere in questione altre regole, o scopi al cui perseguimento è associata la “policy” di seguire la regola senza metterla in questione. Si pensi, ad esempio, alla regola dell'ordinamento militare che prescrive di osservare gli ordini del superiore con prontezza e senza discuterli²¹, e a tipiche giustificazioni di questa forma di rigida obbedienza, come la sua asserita necessità per una coordinazione rapida ed efficace.

Non credo sia necessario argomentare a favore dell'utilità di affrontare il fenomeno del trinceramento non in modo puramente speculativo, ma in stretto rapporto con le scienze cognitive. È solo l'indagine empirica sui rilevanti meccanismi mentali che può risolvere questioni come, ad esempio, se e fino a che

¹⁹ Il lettore filosofo del diritto avrà riconosciuto qui la posizione di Raz. Cfr. J. Raz, “Authority and Justification”, in *Philosophy & Public Affairs*, 14 (1985), n. 1, pp. 3-29, p. 7.

²⁰ Se il trinceramento fosse inteso in senso molto lato, esso potrebbe abbracciare anche le strategie decisionali usualmente indicate come *precommitments*. Qui intendo però il trinceramento in un senso più stretto, che include soltanto dispositivi cognitivi che operano aumentando o proteggendo la forza della regola *in quanto ragione d'azione*. Molti dei tipici *precommitments* non soddisfano questa definizione. Sui *precommitments* v. Th. Schelling, “An Essay on Bargaining”, in *The American Economic Review*, 46 (1956), n. 3, pp. 281-306; J. Elster, *Ulysses and the Sirens. Studies in Rationality and Irrationality*, Cambridge University Press, Cambridge, 1984.

²¹ Art. 729 del DPR n.90/2010.

punto sia possibile mettere in questione una regola “off-line”, senza incidere sul processo decisionale effettivo. E, come sempre, l’indagine empirica sui meccanismi può rilevare aspetti del fenomeno meno evidenti all’introspezione, fino, forse, a suggerirne una riconfigurazione – lo vedremo riguardo al ruolo giocato da dispositivi cognitivi di carattere non raziocinativo ma abituale.

6. Controllo cognitivo e automaticità

L’idea di ragione d’azione (e delle regole come ragioni d’azione), come abbiamo visto, è indissolubilmente connessa alla nozione di controllo intenzionale dell’azione: la capacità di superare impulsi e spinte prepotenti, di focalizzare la riflessione su come sia opportuno agire in base a certi standard, e di mettere effettivamente in atto la scelta a cui si è pervenuti. L’esperienza del controllo contrasta con l’esperienza di azioni o pensieri che, invece, si innescano o procedono da soli, in modo “automatico”, in assenza, e spesso a dispetto, dei nostri tentativi di inibirli o guidarli.

A partire dalla seconda metà del secolo scorso, la nozione di controllo, e il contrasto fra controllo e automaticità, hanno assunto un rilievo centrale nella ricerca di ambito psicologico (soprattutto di psicologia cognitiva e di psicologia sociale) e nelle neuroscienze cognitive. I fenomeni del controllo e dell’automaticità sono stati scomposti, analizzati in termini di meccanismi, e riconfigurati in modi che, almeno in parte, divergono da assunti standard del senso comune e dei suoi prolungamenti filosofici. Si tratta di una ricerca complessa, frastagliata in diversi filoni, in cui si mescolano termini, concetti e modelli diversi, non chiaramente raccordati gli uni agli altri²². In questo paragrafo presenterò, in modo semplificato, alcune linee di ricerca molto influenti.

La prima fase della spiegazione meccanicista è, si ricorderà, la specificazione e scomposizione del fenomeno. Ciò è avvenuto, nel nostro caso, attraverso l’identificazione di una lista di caratteristiche osservabili che distinguono casi paradigmatici di azioni automatiche e controllate a livello introspettivo e comportamentale (come l’azione è esperita dall’agente e come appare all’osservatore esterno)²³. Fra le molte caratteristiche individuate, quelle centrali sono le seguenti²⁴.

²² Per avere un’idea della varietà degli approcci, v. T. Egner (a cura di), *The Wiley Handbook of Cognitive Control*, Wiley, Malden, 2017 e St.B.T. Evans, K. Frankish, *In Two Minds. Dual Process Theories and Beyond*, Oxford University Press, Oxford, 2009.

²³ Per la concettualizzazione di questo tipo di operazione, con riguardo alla nozione di automaticità, v. A. Moors, “Automaticity: Componential, Causal, and Mechanistic Explanations”, in *Annual Review of Psychology*, 67 (2016), n. 5, pp. 5.1-5.25.

²⁴ Il lettore noterà subito la corrispondenza tra le caratteristiche di automaticità e controllo e quelle del Sistema 1 e Sistema 2 in D. Kahneman, *Thinking, Fast and Slow*, Penguin, London, 2011. Quella di Kahneman è solo una delle tante versioni di questa diffusissima contrapposizione.

L'azione automatica è veloce, richiede poca o nessuna attenzione, è inconscia, involontaria, poco flessibile. Dato un certo stimolo, la risposta comportamentale è selezionata e attivata in modo molto rapido, senza o con poco impiego di attenzione, e in modo inconscio – in particolare, senza l'esperienza di un confronto esplicito fra alternative e della scelta deliberata di una di esse. L'azione automatica, inoltre, segue pattern largamente predeterminati: certi tipi di stimoli attivano invariabilmente certi tipi di risposte. I casi paradigmatici di azione automatica sono, ovviamente, reazioni impulsive di carattere emotivo, come paura e fuga, e comportamenti abituali, come fermarsi davanti al semaforo rosso.

Le azioni controllate sono più lente, richiedono più attenzione, e sono accompagnate da un ampio ventaglio di esperienze soggettive – la rappresentazione conscia di più alternative e delle loro conseguenze, sentimenti negativi o positivi verso di esse, senso di scelta volontaria, senso di sforzo mentale e di concentrazione. Infine, le azioni controllate sono più flessibili. La flessibilità è il tratto centrale della più influente ridefinizione del controllo, la nozione di “controllo cognitivo”, inteso come la capacità di orientare flessibilmente l'azione sulla base di uno standard internamente rappresentato (uno scopo, una regola), inibendo risposte più automatiche²⁵.

Il passo successivo è la specificazione, prima in termini funzionali omuncolari e poi in termini neurali, dei meccanismi responsabili delle modalità di azione automatica e controllata, e delle loro tipiche differenze. Si tratta di un lavoro estremamente complesso, e ancora largamente incompleto. Ma alcuni punti importanti sembrano ormai acquisiti.

L'idea di base è che i meccanismi mentali siano organizzati in livelli. Le funzioni cognitive “superiori” sono abilitate da meccanismi di alto livello, che si innestano su, e regolano *top-down*, meccanismi preesistenti di livello più basso. Ma, più che gerarchica, l'organizzazione dei meccanismi mentali è “eterarchica”²⁶: i meccanismi superiori non soltanto regolano, ma sono anche regolati *bottom-up* da, i meccanismi inferiori, in assenza di un coordinamento centrale. L'organizzazione in livelli non è soltanto funzionale, ma anche anatomica. I meccanismi di livello più alto coinvolgono circuiti della corteccia cerebrale, e in particolare della corteccia prefrontale, mentre i meccanismi di livello più basso sono distribuiti in parti più posteriori della corteccia e in aree sub-corticali.

Questa immagine si applica anche al controllo cognitivo, una tipica funzione superiore. Mentre i comportamenti automatici sono orchestrati, in maniera prevalente, da un insieme di meccanismi di livello inferiore, il controllo richiede l'intervento di meccanismi di livello superiore, che coinvolgono la corteccia

²⁵ La letteratura sul controllo cognitivo è vastissima. Un'ottima introduzione si trova in D. Badre, *On Task. How Our Brain Gets Things Done*, Princeton University Press, Princeton, 2020. V. anche J.D. Cohen, “Cognitive Control: Core Constructs and Current Considerations”, in T. Egner, *op. cit.*, pp. 3-28.

²⁶ W. Bechtel, L.T.-L. Huang, *Philosophy of Neuroscience*, Cambridge University Press, New York, 2022, pp. 73-76.

prefrontale. Questi ultimi non causano l'azione direttamente, ma influenzano l'attività di circuiti di livello più basso, responsabili della selezione finale dell'azione²⁷. I meccanismi di livello superiore, inoltre, non si regolano da soli – né sono diretti da un soggetto *dominus* della macchina cerebrale. Sono invece regolati *bottom-up* da meccanismi di livello inferiore, anche qui, sottolineo, in assenza di un coordinamento centrale. L'esperienza del controllo intenzionale di livello personale (*io* decido di agire) emerge attraverso, ed è interamente costituita da, l'interazione di una rete di meccanismi sub-personali.

Fra i meccanismi di alto livello responsabili del controllo, un ruolo centrale è giocato dalla memoria di lavoro, metaforicamente descritta come uno spazio nel quale una quantità limitata di informazione è mantenuta attiva per guidare contestualmente l'azione, o altri processi cognitivi. Secondo un modello molto influente, la memoria di lavoro è realizzata da pattern di attivazione prolungata di insiemi di neuroni distribuiti in varie parti del cervello, tra cui, soprattutto, porzioni della corteccia prefrontale.

I meccanismi di alto livello responsabili del controllo possono essere attivati in gradi e combinazioni diverse, correlate con gradi e combinazioni diverse dei tipici contrassegni osservabili del controllo. Anziché un contrasto netto, la distinzione tra automaticità e controllo va dunque pensata come un continuo graduale e multidimensionale, che va da azioni paradigmaticamente automatiche (in cui i meccanismi del controllo non operano o operano in modo molto ridotto) verso forme di azione sempre più controllata. Tornerò più avanti su alcuni aspetti di questa gradualità.

7. Struttura del controllo

Un episodio paradigmatico di controllo coinvolge le seguenti operazioni.

Il comportamento attivo viene provvisoriamente inibito²⁸, mentre l'attenzione è focalizzata sulla scelta di un corso d'azione.

Viene intrapresa una ricerca che esplora diversi sistemi di memoria per individuare, o costruire, opzioni d'azione ulteriori rispetto a quelle che tendono ad attivarsi automaticamente, simulando la loro esecuzione e immaginando i loro esiti – una forma di pensiero controfattuale, chiamata *episodic future thought*²⁹.

Le opzioni prese in considerazione vengono valutate e selezionate, sotto la guida di “parametri del controllo”. I parametri del controllo sono rappresentazioni “esplicite”, e cioè accessibili alla cognizione di alto livello. Possono essere

²⁷ E.K. Miller, J.D. Cohen, “An Integrative Theory of Prefrontal Cortex”, in *Annual Review of Neuroscience*, 24 (2001), 167-202.

²⁸ Y. Munakata, S.A. Herd, C.H. Chatham, B.E. Depue, M.T. Banich, R.C. O'Reilly, “A Unified Framework for Inhibitory Control”, in *Trends in Cognitive Science*, 15 (2011), n. 10, pp. 453-459.

²⁹ D.L. Schachter, R.G. Benoit, K.K. Szpunar, “Episodic Future Thinking: Mechanisms and Functions”, in *Current Opinion in Behavioral Science*, 17 (2017), 41-50.

conservate nella memoria dichiarativa (e, spesso, riportate verbalmente), e, soprattutto, possono essere attivate nella memoria di lavoro. È attraverso l'attivazione nella memoria di lavoro che i parametri del controllo guidano la valutazione e selezione dell'azione. I parametri del controllo a cui si fa abitualmente riferimento hanno nomi e ruoli funzionali analoghi a figure familiari della psicologia di senso comune: “valori”, “scopi”, “regole”.

Anche nel più semplice episodio di controllo, la valutazione e selezione dell'azione sono svolte attraverso l'interazione di una rete estremamente complessa di circuiti. Per rappresentare questa complessità in modo semplice e vicino al senso comune, ma non biologicamente infondato, è utile distinguere fra circuiti sulla base dei parametri sottostanti: un circuito più “emotivo”, basato su valori (*value-based decision making*), e circuiti più “cognitivi”, basati su scopi o su regole (*goal-based e rule-based decision making*).

Nel circuito basato sul valore, la simulazione delle opzioni prese in considerazione e la anticipazione di loro esiti stimola una cascata di segnali affettivi e motivazionali ad essi associati. Questi segnali esercitano un effetto doppio. Da un lato, stimolano in modo automatico azioni appetitive o avversive. D'altro lato, stimolano un processo di valutazione più complesso, in cui gioca un ruolo centrale la corteccia orbito-frontale. Qui, i segnali evocati dai processi immaginativi sono progressivamente integrati in rappresentazioni esplicite del valore, soprattutto in termini di impatto edonico atteso, attribuito all'esito di un'azione, scontato sulla base della sua probabilità e distanza temporale³⁰. Durante il processo di valutazione, rappresentazioni delle opzioni prese in considerazione sono mantenute attive in aree più posteriori del cervello. Quando le rappresentazioni del valore – o “valori” per semplicità – entrano nella memoria di lavoro, inviano segnali che aumentano il livello di attivazione dell'opzione corrispondente, fino al raggiungimento di una soglia critica di “preferenza”³¹, che può innescare immediatamente l'azione, ovvero essere archiviata in memoria per proseguire la valutazione di altre opzioni.

Gli “scopi” sono, invece, rappresentazioni esplicite di stati di cose che, se mantenuti attivi nella memoria di lavoro, orientano la ricerca verso le opzioni associate al loro ottenimento, e influenzano la selezione dell'azione aumentando il livello di attivazione delle opzioni prese in considerazione sulla base della loro probabilità di successo. Le “regole” sono, invece, rappresentazioni esplicite che associano certi contesti o condizioni a certe risposte (ad es., “Se C, allora A”). La regola è applicabile quando il contesto attuale soddisfa il contesto rappresentato. Una regola applicabile, se mantenuta nella memoria di lavoro, influenza la

³⁰ Un importante modello di questo processo è in A. Bechara, A. Damasio, “The Somatic Marker Hypothesis: A Neural Theory of Economic Decision”, in *Games and Economic Behavior*, 52 (2005), 336-372. Un modello diverso, oggi dominante, fa leva sull'integrazione di segnali di *reward* piuttosto che di markers somatici. V. per tutti J. Wallis, “Orbitofrontal Cortex and Its Contribution to Decision-Making”, in *Annual Review of Neuroscience*, 30 (2007), pp. 31-56.

³¹ Cfr. A. Rangel, T. Hare “Neural Computations Associated with Goal-directed Choice”, in *Current Opinion in Neurobiology*, 20 (2010), pp. 262-270.

selezione dell'azione, aumentando il livello di attivazione dell'opzione rappresentata (nell'esempio, A)³².

Il controllo dell'azione avviene attraverso l'interazione di questi circuiti, a volte fluidamente coordinati (valutiamo positivamente un certo stato di cose, formiamo lo scopo di ottenerlo, e seguiamo una regola associata al suo ottenimento), altre volte in competizione (una regola ci chiede di fare qualcosa a cui attribuiamo un valore negativo). Il livello di attivazione delle opzioni rappresentate, oltre a essere modulato *top-down* dai parametri del controllo, è modulato anche *bottom-up* da circuiti di livello inferiore, che operano in modo automatico. L'azione finale dipenderà da tutte queste complesse interazioni, i cui esiti, estremamente vari, si riflettono nella ricchissima variabilità della nostra esperienza del controllo. A volte, la ricerca e valutazione di opzioni alternative – sostenuta da uno sforzo mirato di concentrazione e, magari, dall'intenzione esplicita di seguire certe procedure decisionali – si sviluppa in modo esteso e accurato, generando strutture di pensiero cosciente che hanno una certa somiglianza con gli alberi decisionali e le computazioni ipotizzate dalla teoria della decisione razionale. Altre volte, la decisione si risolve attraverso brevi flash di immagini sensoriali ed emotive che si susseguono in modo semi-automatico. A volte, percorriamo consciamente gerarchie complesse di scopi e regole interconnesse. Altre volte, ciò che esperiamo è solo il pensiero di una regola molto semplice, e la sua immediata, semi-automatica esecuzione. Secondo molti, inoltre, il controllo può avvenire anche in forma inconscia³³. Anche il livello di granularità e affidabilità della nostra coscienza è molto vario. Così, può accaderci di sapere che qualcosa non ci piace, ma di non essere in grado di specificare quale sia l'aspetto che non ci piace. E può accaderci di interpretare in modo sbagliato il nostro processo di controllo, credendo di essere guidati da parametri che, invece, non hanno giocato alcun ruolo causale³⁴.

8. Controllo cognitivo e ragioni

C'è una evidente corrispondenza tra la analisi meccanicista del controllo cognitivo e

³² Per un quadro generale delle prospettive delle neuroscienze cognitive sulle regole, v. S. Bunge, J. Wallis (a cura di), *Neuroscience of Rule-Guided Behavior*, Oxford University Press, Oxford, 2020. Per un'ottima review della influente linea di ricerca a cui mi rifarò nei paragrafi successivi v. D. Noelle, "On the Neural Basis of Rule-Guided Behavior", in *Journal of Integrative Neuroscience*, 11 (2012), n. 4, pp. 453-475.

³³ C.L. Suhler, P.S. Churchland, "Control: Conscious or Otherwise", in *Trends in Cognitive Science*, 13 (2009), n. 8, pp. 341-347.

³⁴ Su questo conturbante fenomeno v. R.E. Nisbett, T.D. Wilson, "Telling More Than We Can. Verbal Report on Mental Processes", in *Psychological Review*, 84 (1977), n. 3, pp. 231-259; J. Haidt "The Emotional Dog and Its Rational Tail. A Social Intuitionist Approach to Moral Judgment", in *Psychological Review*, 108 (2001), n. 4, pp. 814-834; P. Carruthers, *The Opacity of Mind. An Integrative Theory of Self-Knowledge*, Oxford University Press, Oxford, 2011.

la rappresentazione di senso comune e filosofica del controllo intenzionale dell'azione e del ruolo delle ragioni.

I parametri del controllo – valori, scopi, regole – corrispondono, e devono il loro nome, a figure della psicologia di senso comune che svolgono tipicamente il ruolo di *ragioni operative*. L'interazione dinamica, integrazione e competizione di circuiti diversi, e spesso parametri diversi, fino all'emersione di un segnale dominante a favore di una opzione – una *decisione* – in cui consiste un episodio di controllo corrisponde, se svolto in modo almeno minimamente conscio, alla riflessione o *deliberazione pratica*. Episodi di controllo più estesi, strutturati, sostenuti da sforzo di attenzione, e basati su criteri espliciti, corrispondono al *ragionamento*, nelle sue varie forme. Episodi di controllo minimamente consci e semi-automatici corrispondono alla *intuizione*. La competizione conscia tra parametri di controllo – ragioni – incompatibili, e la sua progressiva soluzione, corrisponde a ciò che concettualizziamo ed esperiamo come *bilanciamento*.

Questa corrispondenza, va sottolineato ancora, è molto parziale. La realtà psicologica dei processi che il senso comune e il gergo filosofico concettualizzano attraverso il lessico delle ragioni ha molti strati complessi. Ma è verosimile che il controllo cognitivo (o un certo grado di controllo cognitivo) ne costituisca lo strato fondamentale. Una ragione, dunque, considerata nel suo ruolo di guida dell'azione, è anzitutto (anche se non solo) una rappresentazione esplicita – valore, scopo o regola – che influenza il processo di riflessione (lo svolgimento dell'episodio di controllo) e la decisione (la selezione dell'azione) al modo di un parametro di controllo: se attivata nella memoria di lavoro, cioè, essa esercita influenza *top-down* sui circuiti che innescano il comportamento, rendendo possibile discostarsi da comportamenti più automatici – risposte impulsive, abitudinarie, o spinte più immediate e prepotenti. Non qualsiasi causa dell'azione, ma una causa specifica, che opera attraverso un processo specifico.

Ma che dire invece dell'aspetto normativo delle ragioni? Può essere accomodato in termini di controllo? Ritengo di sì. Le menti umane sono capaci di applicare il controllo in modo ricorsivo, cioè, di sottoporre processi di controllo a controllo ulteriore. Chiamiamolo “meta-controllo”. L'idea, nel suo nucleo essenziale, è molto semplice. Quando X riconosce consapevolmente che una certa azione A è connessa con P *in modo appropriato*, e che è *corretto* accettare P come ragione, quello che sta facendo è (i) istanziare, o simulare, un processo di controllo nel quale il parametro P orienta verso A; (ii) sottoporre il pattern decisionale istanziato o simulato ad una operazione di meta-controllo, che lo conferma. Il meta-controllo può essere guidato da parametri di controllo di più alto livello – ragioni in virtù delle quali è corretto accettare P come ragione, o schemi di inferenza espliciti. Ma il meta-controllo può anche consistere, più semplicemente, in un loop nel quale vengono riapplicati gli stessi parametri del processo di controllo target, confermandone l'esito – come quando, per controllare un'operazione, la rifacciamo applicando lo stesso algoritmo. Proprio come il controllo, il meta-controllo può assumere forme raziocinative (confermiamo o rigettiamo un certo pattern decisionale sulla base di ragioni esplicite) o intuitive (“sentiamo” che un certo

pattern decisionale è corretto o scorretto). E può anche essere sostituito da una mera aspettativa: mi aspetto che confermerò un pattern decisionale in un processo di meta-controllo. Così come può sussistere soltanto in forma disposizionale: il controllo si svolge ad un solo livello, ma il meta-controllo può attivarsi immediatamente, date certe sollecitazioni (un dubbio o, più tipicamente, una sollecitazione sociale, una critica – ma ho stabilito di lasciare da parte l’aspetto sociale delle ragioni).

È questa, mi sembra, la dimensione normativa, in un senso soggettivo, delle ragioni: P è, per X, una *buona* ragione a favore di A, se è un parametro di controllo che orienta verso A (o che mi aspetto orienterebbe verso A) e se è confermato (o mi aspetto che verrebbe confermato) in un ciclo di meta-controllo. In questo senso, una ragione è sì una causa, ma una causa *buona, corretta* – e cioè, confermata in un nuovo processo causale³⁵.

9. Regole e abitudini nel cervello

Veniamo adesso al secondo aspetto del fenomeno delle regole, la relazione fra regole e abitudini.

Si ricordi che le regole – nei termini del modello di spiegazione meccanicista che sto adottando – sono rappresentazioni mentali esplicite che associano ad un certo contesto o condizione una certa risposta. Nel caso più semplice, viene associato un certo stimolo sensoriale ad una certa risposta motoria (“Se luce rossa, premere il pulsante”). Che una regola sia “esplicita”, si ricordi, significa che può essere conservata nella memoria dichiarativa e attivata nella memoria di lavoro. In quest’ultimo caso, la regola invia segnali che influenzano le parti del cervello che selezionano l’azione, aumentando la probabilità che venga selezionata l’opzione rappresentata.

Le regole, così intese, sono ben distinte da altri tipi di associazioni stimolo-risposta, e in particolare dalle abitudini. Come le regole, le abitudini possono essere descritte da enunciati condizionali come “Se luce rossa, premere il pulsante”. Nel caso delle abitudini, però, ciò che l’enunciato descrive non è una rappresentazione esplicita, ma l’esistenza, nella memoria implicita, di un meccanismo che risponde automaticamente alla detezione di una luce rossa stimolando l’azione del premere il pulsante. Questo meccanismo coinvolge circuiti di livello più basso rispetto a quelli che codificano le regole e dotati di diverse proprietà funzionali. Le diverse proprietà funzionali dei rispettivi circuiti sono responsabili dei tratti controllati e automatici del comportamento guidato da regole e del comportamento abituale. Ripercorriamo rapidamente questi tratti, aggiungendo qualche considerazione riguardo alle diverse modalità di apprendimento di regole e abitudini.

³⁵ Mi permetto di rinviare, su questo punto, a M. Brigaglia, “Genealogia della normatività. La normatività come controllo”, in *Diritto & Questioni pubbliche*, 18 (2018), n. 1, pp. 59-103.

Le abitudini sono apprese lentamente, attraverso ripetuto rinforzo di certi pattern di risposta. Semplificando molto, quando l'adozione di una certa risposta a seguito di un certo stimolo produce una ricompensa (*reward*) maggiore dell'attesa, vi è rilascio di dopamina dal mesencefalo, che ha come effetto il rafforzamento, in una parte dei gangli della base, lo striato, delle connessioni neurali che controllano l'adozione di quella risposta a seguito di quello stimolo, rendendola più probabile³⁶. Una volta apprese, però, le abitudini si perdono con difficoltà, e tendono invariabilmente ad attivarsi in presenza dello stimolo appropriato. Esse, inoltre, dipendono da un meccanismo estremamente affidabile: in assenza di attivazione di controllo inibitorio, il comportamento abituale, una volta innescato, tende con altissima probabilità ad essere eseguito, e ad esserlo in tempi veloci. La stabilità, affidabilità e velocità delle abitudini sono senza dubbio un vantaggio *in ambienti stabili*. Le abitudini, infatti, in virtù delle loro modalità di apprendimento, ripetono comportamenti che, nelle condizioni in cui l'apprendimento è avvenuto, hanno prodotto esiti positivi. Ma quelle stesse caratteristiche si trasformano in uno svantaggio quando il contesto muta – e, come sappiamo bene, il contesto muta continuamente, e i comportamenti abituali hanno spesso esiti nefasti.

Le regole, invece, possono essere apprese molto rapidamente – a volte, come nel caso delle istruzioni, basta un singolo atto comunicativo. Inoltre, come vedremo meglio nel prossimo paragrafo, possono essere rapidamente aggiornate (*updated*) – quando una regola non risulta più appropriata al contesto, possiamo abbandonarla e seguire un criterio di decisione diverso. Ciò dà al comportamento governato da regole una flessibilità che contrasta con la stabilità delle abitudini. La flessibilità delle regole è chiaramente un vantaggio a fronte di contesti mutevoli. Ad essa si accompagnano però svantaggi: una maggiore lentezza del processo decisionale, e una minore stabilità e affidabilità. Le regole, infatti, possono essere perse (dimenticate) più facilmente di una abitudine, sono maggiormente soggette a distrazioni (siamo impegnati a svolgere un compito sulla base di una regola, ma ci distraiamo, e non svolgiamo il compito o non seguiamo la regola) e spesso, seppur attive, non riescono a determinare il comportamento – ad esempio, perché soverchiate da impulsi contrastanti. Queste caratteristiche delle regole dipendono dalle proprietà funzionali dei relativi circuiti neurali – dei circuiti che costituiscono la memoria dichiarativa nella quale sono conservate le regole, dai circuiti che costituiscono la memoria di lavoro in cui le regole sono attivate, dai circuiti che connettono la memoria di lavoro alla selezione dell'azione, e dai circuiti che regolano l'aggiornamento o il mantenimento dei contenuti della memoria di lavoro.

Fra il comportamento guidato da regole e il comportamento abituale vi sono però non soltanto rilevanti differenze, ma anche importanti aspetti di *continuità*. Nel prossimo paragrafo, mi soffermerò sui tratti abituali dei meccanismi che

³⁶ H.H. Yin, B.J. Knowlton, "The Role of Basal Ganglia in Habit Formation", *Nature Reviews. Neuroscience*, 7 (2006), pp. 465-476.

regolano l'attivazione delle regole nella memoria di lavoro. Qui, invece, dirò qualcosa sulla automatizzazione del comportamento.

La regola, si ricorderà, non causa direttamente il comportamento, ma invia segnali *top-down* che aumentano il livello di attivazione di una rappresentazione dell'opzione corrispondente, codificata in altre parti del cervello (come la corteccia premotoria), fino al raggiungimento di una soglia critica, data la quale l'opzione è selezionata. L'attivazione dell'opzione corrispondente dipende anche da segnali *bottom-up*. Quanto più forti sono i segnali *bottom-up* convergenti con la regola, tanto minore sarà l'apporto ad essa richiesto. Durante la formazione dell'abitudine, che può avvenire anche attraverso la reiterazione del comportamento sotto la guida della regola, la forza dei segnali *bottom-up* aumenta, e la responsabilità della regola decresce. Quando l'abitudine si sarà consolidata, la regola diventerà superflua, e, pur se ancora iscritta nella memoria dichiarativa, resterà inattiva. In questa situazione, continuiamo ad agire secondo il pattern indicato dalla regola, non a causa della regola, ma in via di abitudine. Se, però, la regola continua ad esistere nella memoria dichiarativa e, se attivata, confermerebbe il comportamento, e se vi sono meccanismi affidabili che riattiverebbero la regola in caso di deviazione dal corso abituale, si può sensatamente dire che il comportamento, seppur abituale, è "sensibile" alla regola, o implicitamente regolato da essa³⁷. Si pensi alla distanza conversazionale: la rispettiamo in modo automatico, in via di abitudine; ma quando viene violata dall'interlocutore o abbiamo noi stessi la tentazione di violarla, si attiva automaticamente una regola che ci induce a ripristinare la distanza.

10. Regole trincerate e meccanismi di *gating*

Affinché la regola possa influenzare efficacemente la selezione del comportamento, è necessario che (i) il segnale da essa inviato abbia forza sufficiente (e quale sia la soglia di sufficienza dipende anche dalla forza di segnali contrastanti); (ii) che la regola continui a inviare segnali, e cioè che si mantenga attiva nella memoria di lavoro.

Possiamo chiamare "stabilità" della regola il suo mantenimento nella memoria di lavoro. Come sappiamo fin troppo bene, le regole (e, in genere, i contenuti della memoria di lavoro) hanno una stabilità ridotta. Tendiamo facilmente a distrarci, a spostare l'attenzione su nuovi stimoli. E se una regola "esce di mente" – non è più attiva nella memoria di lavoro –, non potrà guidare la nostra azione.

D'altra parte, una eccessiva stabilità delle regole sarebbe dannosa. Il vantaggio del controllo è proprio la flessibilità, il rapido adattamento a nuovi contesti. Se una regola fosse eccessivamente stabile, continuerebbe a orientare il comportamento anche quando ciò fosse, dato il contesto mutato, del tutto inappropriato. Quello che serve – e quello che bene o male abbiamo – è un *trade-*

³⁷ Cfr. H.L.A. Hart, *op. cit.*, pp. 55-57.

off tra stabilità e flessibilità. L'ipotesi di spiegazione attualmente più accreditata del modo in cui il nostro cervello realizza questo *trade-off* è incentrata sui cosiddetti meccanismi di *gating*, distinti in *input gating* e *output gating* (per brevità, *i-* e *o-gate*)³⁸. I dettagli del funzionamento dei meccanismi di *gating* sono molto complessi, ma l'idea centrale è semplice.

L'*i-gating* regola, in modo selettivo, il mantenimento o aggiornamento dei contenuti della memoria di lavoro. Può essere metaforicamente descritto come la regolazione di un cancello che circonda una specifica informazione attiva nella memoria di lavoro – nel caso per noi più interessante una regola. Quando l'*i-gate* è chiuso, la regola è mantenuta attiva, protetta da segnali che potrebbero interferire con essa. Quando l'*i-gate* è aperto, la regola può essere rimpiazzata da altri contenuti – ciò che tipicamente accade quando, durante un compito, siamo distratti da altre preoccupazioni.

L'*i-gate* è regolato da una struttura sub-corticale parte dei gangli della base, lo striato³⁹. Lo striato opera attraverso un meccanismo del tutto analogo a quello responsabile dei comportamenti volontari automatici e della formazione di abitudini. Per descriverlo in modo semplice, si può ricorrere alla metafora di una "policy" di *i-gating*, e cioè chiudere o aprire l'*i-gate* in una certa circostanza. Quando una *policy* di chiusura o apertura dell'*i-gate* rispetto a una regola R produce una ricompensa inattesa, un rilascio di dopamina rafforza circuiti dello striato che sostengono la stessa *policy*, rendendo più probabile che essa venga attuata nuovamente in futuro. In questo modo, una storia di ricompense promuove la maggiore stabilità o flessibilità della regola⁴⁰.

Questa spiegazione del funzionamento della memoria di lavoro è molto importante. Essa, in primo luogo, elimina l'omuncolo che, nella forma di un esecutivo centrale, affliggeva le concezioni originarie della memoria di lavoro. Al suo posto, abbiamo un meccanismo modellato su quello, relativamente ben compreso, responsabile della formazione delle abitudini. In secondo luogo, modifica, come già accennato, il quadro generale dell'organizzazione del cervello. Al posto di una organizzazione *gerarchica*, in cui meccanismi di alto di livello come la memoria di lavoro regolano *top-down* meccanismi di livello inferiore, abbiamo una organizzazione *eterarchica*, in cui i meccanismi di alto livello sono, a loro volta, regolati *bottom-up* da meccanismi di livello inferiore, come i gangli della base. Infine, riconfigura il fenomeno del rapporto tra regole e abitudini: al cuore del funzionamento delle regole, vi sono meccanismi analoghi alle abitudini!

³⁸ Per una eccellente introduzione v. D. Badre, *op. cit.* Per una review più specialistica v. D. Noelle, *op. cit.*

³⁹ M.J. Frank, B. Loughry, R.C. O'Reilly, "Interactions between the Frontalcortex and Basal Ganglia in Working Memory: A Computational Model", in *Cognitive, Affective, and Behavioral Neuroscience*, 1 (2001), pp. 137-160.

⁴⁰ R.C. O'Reilly, M.J. Frank, "Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia", in *Neural Computation*, 18 (2006), pp. 283-328.

L'aspetto sul quale vorrei adesso concentrarmi, però, è il ruolo che i meccanismi di *i-gating* possono avere nel trinceramento. Il trinceramento, si ricorderà, consiste in dispositivi cognitivi che proteggono la regola contro la sua messa in questione e riconsiderazione. Il trinceramento di una regola può consistere, almeno in parte, in dispositivi raziocinativi – come regole che indicano di non mettere in questione o riconsiderare altre regole. Ma simili dispositivi raziocinativi sono soggetti agli stessi limiti delle regole che dovrebbero proteggere – sono efficaci solo nella misura in cui sono mantenuti attivi nella memoria di lavoro. Sarebbe molto più utile un dispositivo non soggetto a questi limiti. E questo dispositivo esiste: è una *policy* che mantiene chiuso l'*i-gate* rispetto ad una certa regola. La coincidenza delle metafore è indicativa: un *i-gate* chiuso è come una trincea, che protegge la regola contro interferenze che potrebbero scazarla. Mi sembra estremamente plausibile che il dispositivo che, in via primaria, trincera una regola R (o una classe di regole R), sia costituito, almeno in parte, da una *policy* di chiusura dell'*i-gate* rispetto a R (magari insieme a meccanismi che inibiscono la messa in questione di R).

I meccanismi di *i-gating*, si noti, non sono di tipo raziocinativo. Non sono guidati da ragioni. Piuttosto, sono simili ad abitudini. Sono plasmati lentamente attraverso una storia di apprendimento implicito, inconscio, guidato da ricompense. E come le abitudini, le *policies* di *i-gating* rispetto ad una regola o classe di regole R possono essere addestrate attraverso il ricorso mirato a ricompense (e punizioni). Se viene ricompensata una *policy* di chiusura, cioè trincererà la regola, riducendo la probabilità che essa venga messa in questione. Un esempio ovvio è l'addestramento dei militari.

Uno dei problemi principali del trinceramento è la sua compatibilità con una applicazione *ragionevole* delle regole. Si tratta, ancora una volta, di un *trade-off* tra stabilità e flessibilità. Può essere auspicabile che certe regole siano trincerate, per aumentare la probabilità che vengano osservate (stabilità). Ma, allo stesso tempo, non è auspicabile che esse tendano a essere osservate anche quando è del tutto irragionevole, sconsigliato farlo (flessibilità). Anche qui, i meccanismi di *i-gating* potrebbero giocare un ruolo importante. Una volta che siano state stabilmente adottate *policies* di chiusura dell'*i-gate* rispetto ad una regola o classe di regole R, è plausibile che si possano anche addestrare, attraverso ricompense (lodi, approvazione, premi), *policies* di apertura dell'*i-gate* in certe circostanze eccezionali (ad esempio, di fronte di una vivida intuizione di scorrettezza dell'azione indicata da R), che facilitano la messa in questione di R.

Veniamo infine al problema se sia possibile mettere in questione una regola "off-line", senza riconsiderarla. Sembrerebbe che mettere in questione una regola tenda irresistibilmente a stimolare la sua riconsiderazione. Se si indugia nel pensiero di ragioni che confliggono con la regola, queste ragioni, sembrerebbe, vengono attivate nella memoria di lavoro, generando una cascata di effetti, compresi segnali che influenzano la selezione dell'azione. A meno che non sia possibile intrattenere "off-line" queste ragioni, pensare ad esse tenendole, per così dire, in quarantena. Qui potrebbe entrare in gioco un *o-gate*. L'*o-gate* regola il

flusso di segnali attraverso i quali una rappresentazione attiva nella memoria di lavoro influenza *top-down* la selezione dell'azione⁴¹. Se rispetto a una certa rappresentazione l'*o-gate* viene chiuso, quella rappresentazione, sebbene attiva, non può influenzare l'azione – è mantenuta “off-line”. Se fosse possibile apprendere una *policy* di chiusura dell'*o-gate* rispetto a certe ragioni R1 che potrebbero militare contro una regola R, a sua volta trincerata attraverso una *policy* di chiusura dell'*i-gate*, si otterrebbe qualcosa di molto simile al processo di esclusione ipotizzato da Raz: le ragioni R1 *o-gated* potrebbero pur sempre essere usate per mettere in questione R, ma si tratterebbe di una messa in questione “off-line”, che si svolge senza influire sulla selezione del comportamento, che resta guidato da R.

⁴¹ C.H. Chatham, M.J. Frank, D. Badre, “Corticostriatal Output Gating during Selection from Working Memory”, in *Neuron*, 81 (2014), 930-942.