

Studi

Mindreading, introspezione e metacognizione: implicazioni per la neuropsichiatria cognitiva

Rossella Guerini e Massimo Marraffa

Ricevuto il 15 novembre 2012, accettato il 23 giugno 2013

Riassunto In questo articolo ci proponiamo di portare alla luce i netti confini che separano i concetti di *mindreading*, introspezione e metacognizione con l'obiettivo di dissipare alcuni fraintendimenti presenti nella letteratura clinica. A tal fine, iniziamo identificando due posizioni principali nell'odierno dibattito filosofico-cognitivo sull'introspezione: da un lato le teorie che sostengono che "introspezione" è la denominazione impropria per un processo *interpretativo*; da un altro lato le teorie che continuano a ritenere che almeno in alcuni casi l'accesso alla propria mente sia *diretto e non interpretativo*. Dopodiché prendiamo posizione contro le teorie dell'accesso diretto e in favore di una certa versione dell'approccio interpretativista. Infine, da questo approccio ricaviamo alcune linee guida per l'uso della *Theory of Mind* in neuropsichiatria cognitiva.
PAROLE CHIAVE: Introspezione; Mindreading; Confabulazione; Metacognizione; Psicopatologia.

Abstract *Mindreading, Introspection, and Metacognition: Implications for Cognitive Neuropsychiatry* – The aim of this article is to draw sharp boundaries between the concepts of *mindreading*, introspection and metacognition in order to rectify some common misunderstandings in the clinical literature. To this aim, we begin with identifying two main approaches in the current philosophical and psychological debate on introspection: at one end of the spectrum, the view that "introspection" is a misnomer for an *interpretative* activity; at the other end of the spectrum, the view that, at least in some cases, the access to one's own inner life is *direct and non-interpretative*. Thereafter, we side against "direct access" theories and in favor of a particular version of the interpretativist approach to introspection. Finally, we propose some guidelines based on this approach for applying a Theory of Mind in cognitive neuropsychiatry.

KEYWORDS: Introspection; Mindreading; Confabulation; Metacognition; Psychopathology.



IN QUESTO ARTICOLO CI PROPONIAMO di tracciare confini netti fra i concetti di *mindreading*, introspezione e metacognizione al fine di dissipare alcuni fraintendimenti presenti nella

letteratura clinica. Nella modellistica clinica, infatti, queste nozioni si trovano troppo spesso schiacciate l'una sull'altra, con grave danno per la tenuta delle teorie.

M. Marraffa - Dipartimento di Filosofia - Università di Roma Tre (✉)
E-mail: marraffa@uniroma3.it

R. Guerini - Dipartimento di Psicologia e Scienze Cognitive - Università di Trento
E-mail: rossella.guerini-1@unitn.it



Nel perseguire questa opera di chiarificazione, procederemo nel modo seguente. Dopo aver identificato quelle che, a nostro giudizio, sono le più solide acquisizioni in materia di comprensione della mente altrui, esamineremo criticamente le ricerche sulla comprensione della *propria* mente (= introspezione), giungendo alla conclusione che la posizione più congruente con la letteratura sperimentale è quella che concepisce l'introspezione come una combinazione di accesso sensoriale e interpretazione.

Come vedremo, questo concetto di introspezione va tenuto ben distinto da quello di metacognizione. Infine, forti di questo risultato, rivolgeremo alcuni commenti critici all'uso della *Theory of Mind* in clinica.

■ Mindreading

Fra le differenti ipotesi che nel corso degli ultimi trent'anni gli psicologi hanno proposto circa la natura e lo sviluppo della capacità di mindreading, la più accreditata resta la *Theory Theory* (TT): siamo in grado di fare psicologia ingenua perché possediamo una *teoria* spontanea, o primaria, della mente (una *Theory of Mind*, ToM), ovvero un corpus di conoscenze integrate e coerenti relative al dominio dei fenomeni psicologici. Si tratta di una tesi che è però suscettibile di differenti curvature.

Nel caso di Josef Perner si può parlare di una versione costruttivista-piagetiana della TT.¹ Qui lo sviluppo di un concetto rappresentazionale di credenza trae origine dalla comprensione della natura delle rappresentazioni *in generale*, non specificamente di quelle mentali. Ossia: il bambino prima impara a comprendere le proprietà delle rappresentazioni esterne o pubbliche; e solo in un secondo momento egli estende, per analogia, tali proprietà alle rappresentazioni mentali.

In particolare, il bambino diviene capace di superare i ben noti compiti di falsa credenza (*False Belief*, FB) allorché, fra i 3 e i 4 anni, l'apparato teorico che utilizza per ragionare sulle menti altrui si arricchisce di un concetto rappresentazionale di credenza precedentemente assente: è tale concetto, infatti, che gli consente

di comprendere che, al pari delle rappresentazioni pubbliche, anche le rappresentazioni interne possono descrivere erroneamente stati di cose.

A questa versione piagetiana della TT fa da contraltare la versione modularista-chomskiana di Alan Leslie.² La ToM che rende possibile l'esercizio del *mindreading* è ora immagazzinata in uno o più meccanismi specifici per dominio e innati, che gradualmente divengono operativi nel corso dello sviluppo infantile. E dunque già nel corso del secondo anno il bambino dispone di un *Theory-of-Mind Mechanism* (ToMM), una componente specializzata dell'intelligenza sociale che riceve in input informazioni concernenti le azioni presenti e passate di altri soggetti e su questa base ne computa i probabili stati psicologici. Gli output di ToMM sono descrizioni di stati psicologici che hanno la forma di *metarappresentazioni* specificamente mentalistiche.

Le due versioni della TT appena menzionate fanno dunque previsioni differenti in merito all'ontogenesi della metarappresentazione mentalistica. Perner colloca la sua comparsa fra i 3 e i 4 anni, quando il bambino acquisisce un nuovo concetto rappresentazionale di credenza. Leslie ritiene invece che ToMM sia già operativo nel secondo anno di età, dal momento che a quell'epoca (in media a 18 mesi) il bambino esercita la sua facoltà metarappresentazionale nel gioco di finzione. In quest'ultimo caso viene però da obiettare: se ToMM è già operativo nel secondo anno di età, perché i test di FB non possono essere superati prima dei 4 anni?

La risposta di Leslie è che, sebbene il concetto di credenza sia già presente in bambini di età inferiore a 4 anni, nei test di FB esso viene mascherato dall'imaturità di un'altra capacità necessaria per la corretta esecuzione del compito: il controllo inibitorio. ToMM impiega come strategia di default quella di attribuire credenze dotate di un contenuto che rispecchia lo stato di cose reale.

Pertanto, per svolgere i compiti di FB è necessario inibire questo automatismo e selezionare un contenuto non fattuale – il compito di un meccanismo di attenzione selettiva (si noti: generale per dominio) denominato *Selection*

Processor (SP). Dunque, i bambini di 3 anni falliscono nei compiti della FB perché posseggono ToMM, ma non ancora SP. Nel caso del gioco di finzione, invece, ToMM non necessita dell'intervento dell'SP dal momento che la modificazione della realtà è decisa dal bambino stesso o da qualcuno che giocando con lui la rende saliente con il suo comportamento.

Il modello ToMM/SP ha cercato conferma in una serie di esperimenti sulla comprensione di rappresentazioni mentali e pubbliche in bambini dallo sviluppo tipico e bambini con *Autistic Spectrum Disorders* (ASD).³ I bambini dallo sviluppo tipico di 3 anni falliscono sia nei compiti di FB standard, sia in due test che misurano la comprensione di rappresentazioni non già mentalistiche ma pubbliche (foto e mappe, rispettivamente). Invece, i bambini ASD forniscono una buona prestazione nei due test non mentalistici ma falliscono nei compiti di FB. I bambini di 4 anni eseguono con successo tutti i compiti.

Il modello ToMM/SP spiega questi dati nel modo seguente: i bambini di 3 anni possiedono ToMM ma non ancora SP; i bambini ASD hanno un ToMM compromesso ma un SP intatto; i bambini di 4 anni possiedono sia ToMM che SP.

Questi risultati rappresentano un bel grattacapo per Perner, il quale – come sappiamo – sostiene che i bambini prima comprendono le rappresentazioni pubbliche, e quindi applicano questa comprensione ai fenomeni mentali. Se così fosse, i bambini ASD dovrebbero trovarsi a mal partito con entrambi i tipi di rappresentazione; e invece la loro prestazione nei compiti sopra citati rivela una dissociazione fra la comprensione di foto datate e di mappe erranee da un lato, e la comprensione di credenze erranee dall'altro.

Il dato segna un punto a favore della prospettiva modularistica di Leslie: i bambini ASD manifestano non già un deficit di comprensione che interessa la rappresentazione in generale, bensì un difetto a carico specificamente della rappresentazione mentale.⁴

Tuttavia una serie di studi che si sono avvalsi di versioni *non verbali* dei compiti di FB

sollevano una grave difficoltà per l'ipotesi del *Selection Processor*. Essi propongono di retrodatare la comparsa della capacità di attribuire credenze erranee dai 4 anni ai 15 mesi circa; e quindi sottoscrivono senz'altro la tesi dell'esistenza di un sistema neurocognitivo che costituisce il nucleo primario innato delle nostre capacità mentalistiche.⁵ A prima vista, questi studi parlano in favore del modello ToMM/SP.⁶

A ben guardare, però, non è così: essi mostrano che intorno alla metà del secondo anno di vita il bambino possiede un meccanismo che è già in grado di selezionare il contenuto della credenza erronea e di generare la conseguente aspettativa circa la condotta dell'agente bersaglio; perché allora tale meccanismo dovrebbe trovarsi in difficoltà quando il test ha natura verbale?

In altre parole, l'ipotesi del *Selection Processor* non spiega lo scarto di due o più anni che intercorre fra il momento in cui il bambino supera i test di mindreading non verbali e il momento in cui riesce a superare quelli verbali.

Si può allora congetturare che sia non già ToMM, ma il linguaggio ad assumere la realtà quale opzione di default; tale strategia sarebbe poi inibita dal graduale incremento della *vigilanza epistemica*.⁷ Quest'ultima è una condizione di efficienza dei sistemi comunicativi. Infatti, la comunicazione è un'attività cooperativa, e come tutte le forme di cooperazione può essere parassitata e minata da *free riders*. Un agente deve perciò premunirsi contro la minaccia dell'inganno, e quindi professare un salutare scetticismo nei riguardi di quanto gli viene comunicato da altri agenti.

Si può allora congetturare che il bambino, oltre a poter contare su un incremento della funzione esecutiva e della memoria di lavoro, divenga sempre più vigile nei confronti dell'eventualità che venga prodotto il falso in contesti conversazionali. E questo spiegherebbe lo scarto di prestazione fra i test di FB non verbali e quelli verbali.⁸

Fin qui abbiamo preso in esame alcuni argomenti in favore dell'esistenza di un sistema multicomponentiale di *mindreading*, il quale assai precocemente (forse già intorno alla metà

del secondo anno) costruisce metarappresentazioni mentalistiche avvalendosi di una base di conoscenze innata e specifica per il dominio della psicologia ingenua. Componente essenziale dell'intelligenza sociale, questo sistema è inserito in un'architettura in cui verosimilmente compaiono altri meccanismi specifici di dominio,⁹ meccanismi generali per dominio (come le funzioni esecutive e la memoria di lavoro) e – come è stato proposto da Alvin Goldman – sistemi in grado di mettere in atto simulazioni mentali di basso e alto livello.¹⁰

Infine, si noti: ToMM è a fondamento di una forma *primaria* di mindreading, la quale va accuratamente distinta dalla sua forma *matura*. Nel corso dello sviluppo le capacità mentalistiche primarie sono cooptate da processi cognitivi superiori al fine di svolgere compiti più complessi. Ne risultano attività mentalistiche avanzate, che interagiscono con altri processi cognitivi (essenzialmente i meccanismi esecutivi, attentivi e progettuali), manifestando una variabilità fra individui e culture.¹¹ Dunque, la sottolineatura dei fondamenti neurocognitivi delle capacità mentalistiche non esclude in alcun modo un'ingente influenza da parte di fattori socio-culturali.

Quale potrebbe essere la funzione adattiva del *mindreading*? Negli ultimi decenni le scienze biologiche e psicologiche hanno chiarito sempre meglio che ogni individuo è portatore di una serie molto complessa di motivazioni, le quali sono sempre e fin dall'inizio *relazionali* e danno luogo a situazioni complesse di compromesso fra due macro-sistemi motivazionali: il primo dedito all'auto-assertività e alla competizione, il secondo volto alla pro-socialità e alla cooperazione. L'essere umano è naturalmente portato alla competizione, e talora alla distruttività, ma anche a forme di socialità, di cooperazione, e perfino di altruismo.¹²

Nel quadro di questa rinnovata antropologia, l'intelligenza umana non può che essere concepita come un'intelligenza primariamente sociale: ovvero una capacità in larga misura dedita all'analisi e alla rielaborazione di rapporti interpersonali complessi.

È più che legittimo supporre, allora, che il

mindreading, in quanto elemento essenziale della nostra intelligenza sociale, si sia evoluto per fornire un vantaggio adattivo sia nel perseguimento delle finalità del sistema competitivo, come è asserito (oramai da più di vent'anni) dai sostenitori della cosiddetta "ipotesi dell'intelligenza machiavellica",¹³ sia in ottemperanza alle richieste del sistema cooperativo, come si è più recentemente proposto.¹⁴

E in effetti, è plausibile ipotizzare un nesso strettissimo fra l'evoluzione di forme di mindreading specificamente umane e la comparsa di sistemi sociali che prevedono un'elevata cooperazione. Michael Tomasello, per esempio, ha individuato una cesura fra il mindreading caratteristico degli esseri umani e quello dei primati proprio in riferimento alla sfera cooperativa.¹⁵ A partire dal secondo anno di vita il bambino manifesta una propensione a impegnarsi in attività cooperative basate sul monitoraggio degli scopi e delle intenzioni altrui (per esempio nel gioco di finzione); gli scimpanzé, invece, sembrano badare alle intenzioni altrui esclusivamente nell'ambito di situazioni competitive.

La teoria della parità io/altro

Durante gli anni '80 e '90 il lavoro in *Theory of Mind* si è occupato quasi esclusivamente dei meccanismi che sono alla base della lettura della mente *altrui*. E tuttavia nell'ultimo decennio un numero crescente di psicologi e filosofi ha proposto teorie dei meccanismi soggiacenti la lettura della *propria* mente. Ciò ha richiesto una sinergia con altre tradizioni di ricerca, prime fra tutte le indagini sulla dissonanza cognitiva e l'attribuzione causale in psicologia sociale, nonché gli studi neuropsicologici sulla confabulazione.¹⁶

Un buon esempio di studio sperimentale sull'introspezione è quello di Petter Johansson e collaboratori, che per analizzare i report introspettivi si sono avvalsi del paradigma cosiddetto della "*Choice Blindness*".¹⁷ Il paradigma consiste in una serie di prove, ciascuna consistente in questo: al partecipante sono presentate le foto di due donne e gli viene chiesto di indicare quale giudica la più interessante; una volta

compiuta la scelta, il partecipante è invitato a dichiarare le ragioni della preferenza.

Alcune prove sono però manipolate: lo sperimentatore/prestigiatore scambia le due immagini e chiede al soggetto, ignaro del trucco, di esplicitare le ragioni della sua preferenza in relazione alla foto che *non* ha scelto (fig. 1a).

I risultati sono stati i seguenti: solo nel 28% delle prove manipolate i partecipanti si sono resi conto dello scambio; nel restante 72% delle prove i soggetti hanno offerto spiegazioni volte a giustificare una scelta che non avevano mai compiuto, spesso menzionando fattori che non potevano aver fatto parte delle ragioni per la scelta. Addirittura, nel 13% dei casi i soggetti hanno giustificato le loro scelte facendo riferimento a caratteristiche del volto non scelto che non erano presenti nel volto scelto!

Per esempio, un partecipante spiegò di aver scelto il volto che gli era stato mostrato dicendo “mi piacciono i suoi orecchini”, quando in realtà la donna nell’immagine scelta non indossava orecchini (cfr. la categoria “confabulazione specifica” in fig. 1b). Questo è un chiaro caso di confabulazione delle motivazioni di una scelta.

Inoltre, dal confronto fra le risposte dei soggetti nella condizione di scelta effettiva e in quella manipolata non è emersa alcuna differenza sistematica. I report introspettivi nelle due condizioni non hanno evidenziato alcuna differenza in ordine al grado di coinvolgimento emotivo, al livello di dettaglio della descrizione e al grado di sicurezza che i partecipanti hanno espresso in merito alla loro scelta.¹⁸

L’assenza di una differenza fra i report manipolati e quelli non manipolati solleva dubbi anche in merito alla genesi dei secondi: la confabulazione – speculano i ricercatori – potrebbe essere la norma e l’esistenza di un report veridico qualcosa da dimostrare.

Questo tipo di dati sperimentali (che potrebbero essere moltiplicati a piacere) ci dicono almeno due cose. Primo, esiste una cognizione inconscia di alto livello;¹⁹ comportamenti, scelte, giudizi possono rispondere a fattori motivazionali non disponibili per l’introspezione e l’esplicitazione dichiarativa. Secondo, abbiamo una normale attitudine alla confabulazione: i

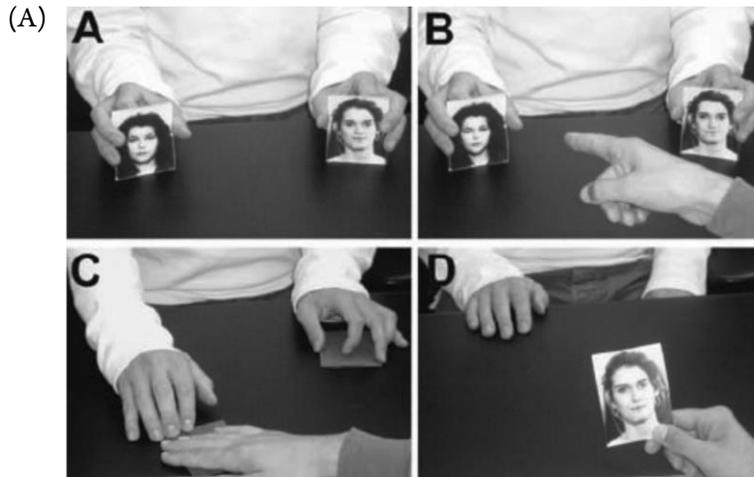
report dei soggetti sperimentali non sono ricostruzioni di processi mentali reali, basate su una consapevolezza introspettiva diretta; sono piuttosto l’esito di un’attività razionalizzante o confabulatoria, in cui il soggetto fabbrica, in base a teorie esplicative socialmente condivise o a una teorizzazione idiosincratica, spiegazioni ragionevoli ma immaginarie della propria condotta.

In questo quadro, allora, il soggetto non gode più di un accesso privilegiato alla propria interiorità. È piuttosto impegnato in un’attività interpretativa basata su meccanismi informati da teorie esplicative che si applicano in egual misura a lui stesso e agli altri. L’input di tali meccanismi è l’osservazione di stati di cose *esterni*, ossia il comportamento del soggetto e la situazione in cui esso agisce; informazioni, dunque, nei cui riguardi il soggetto non gode di alcuna particolare autorità epistemica. È questa una teoria della conoscenza di sé che stabilisce una “parità io/altro”.²⁰

La teoria della parità io/altro ha trovato consenso anche presso alcuni psicologi che spiegano l’ontogenesi del mindreading in base alla TT. Alison Gopnik, ad esempio, ha sostenuto che il *mindreading* in prima persona è reso possibile dalla medesima teoria su cui si basa il *mindreading* in terza persona; e ciò consente di fare la seguente previsione: qualora la teoria non abbia ancora raggiunto il grado di sviluppo richiesto ai fini dello svolgimento di un compito di mindreading in terza persona, essa non consentirà neppure l’espletamento del corrispondente compito in prima persona.²¹

E in effetti, una meta-analisi che riassume i dati provenienti da 178 studi condotti con differenti compiti di FB attesta che nello sviluppo tipico la prestazione nelle varianti in prima e in terza persona del compito è praticamente identica a tutte le età.²² Inoltre, alcuni esperimenti hanno documentato che bambini ASD manifestano il medesimo ritardo nell’esecuzione delle versioni in prima e in terza persona del compito degli Smarties.²³

Gli ASD hanno fornito ulteriori dati a sostegno di una simmetria funzionale fra la comprensione della propria mente e di quella altrui.



(B)

Type	%		
Specific Conf.	13.3		She's radiant. I would rather have approached her at a bar than the other one. I like earrings! [M]
Detailed Conf.	17.3	She looks like an aunt of mine I think, and she seems nicer than the other one. [F]	
Emotional Conf.	9.3		Yes, well, [laughter] she looks very hot in this picture. [M]
Simple Conf.	10.8		Just a nice shape of the face, and the chin. [M]
Relational Conf.	21.3		I thought she had more personality, in a way. She was the most appealing to me. [F]
Uncertainty	11.6	Eh... I don't know. [F]	
Dynamic report	5.2		Oh, [short laughter] Why did I choose her? She looks very masculine! [M]
Original choice	11.2	Because she was smiling. [F]	

Figura 1. (A) Sequenza della procedura di scelta in una prova manipolata. (B) Distribuzione percentuale dei contenuti dei report introspettivi relativi alle prove manipolate, articolata in otto categorie potenzialmente atte a differenziare le prove manipolate da quelle non manipolate. La confabulazione specifica riguarda i report che si riferiscono a caratteristiche specifiche del volto non scelto. La confabulazione dettagliata e quella emozionale sono presenti in report che si collocano molto in alto nella scala relativa a queste due dimensioni (>4.0 su una scala da 1 a 5). La confabulazione semplice e quella relazionale riguardano report in cui la genericità della descrizione del volto impedisce di associarli in modo conclusivo a una delle due foto. Vi sono poi i report dominati dall'incertezza (<2 su una scala da 1 a 5). I report dinamici sono quelli in cui i partecipanti riflettono sulla loro scelta. Infine vi sono i report che si riferiscono al contesto originario della scelta. Le frasi campione sono state tradotte dallo svedese; le lettere (M) e (F) indicano il sesso maschile o femminile del partecipante. Fonte: P. Johansson, L. Hall, S. Sikström, A. Olsson, Failure to Detect Mismatches Between Intention and Out-come in a Simple Decision Task, in: «Science», vol. 310, 2005, pp. 117-118.

Fra questi merita di essere citato almeno lo studio di David Williams e Francesca Happé, che ha indagato la capacità di attribuire le intenzioni sia in terza che in prima persona.²⁴ In un esperimento, i partecipanti (bambini ASD e controlli dallo sviluppo tipico e atipico) erano invitati a completare un disegno su un trasparente (per es. un ragazzo che canta in un coro ed è privo di un orecchio). All'insaputa del bambino, un secondo trasparente su cui compariva un altro disegno incompleto (per es. una tazzina da caffè priva del manico) era stato sovrapposto in cima al primo trasparente.

In tal modo, i partecipanti finivano inintenzionalmente col completare il disegno nella parte superiore del trasparente (ossia il manico della tazzina) piuttosto che il disegno nella parte inferiore (ovvero l'orecchio del ragazzo). Una volta rivelato lo stratagemma, al bambino veniva posta la domanda test: "che cosa avevi intenzione di disegnare?"

In una fase successiva, i bambini guardano un video in cui lo stesso compito era eseguito da un altro bambino; e la stessa domanda test gli fu posta in III persona. Ebbene, in relazione alla capacità di identificare le intenzioni proprie e altrui i bambini autistici hanno fornito una prestazione significativamente peggiore rispetto ai bambini dallo sviluppo atipico (ma "associati" per abilità). E in entrambi i gruppi il successo è risultato fortemente correlato col successo in un certo numero di compiti di FB. Da questi dati sembra possibile arguire che nei bambini autistici la capacità di attribuire intenzioni è deficitaria tanto in prima che in terza persona, e che entrambi i deficit discendono dalle difficoltà che questi bambini hanno con la mentalizzazione in generale.

Alla teoria della parità io/altro si è però giustamente obiettato che nessuna delle sue formulazioni offre una teoria esaustiva dell'autoconoscenza; resta sempre un margine per una forma di conoscenza della propria mente qualitativamente diversa dalla conoscenza delle menti altrui. Per esempio, Nisbett e Wilson tracciano una netta distinzione fra i *processi* causali soggiacenti i giudizi, le decisioni, le emozioni e le sensazioni e il *contenuto* mentale,

ovvero i giudizi, le decisioni, le emozioni e le sensazioni stesse. A questo contenuto il soggetto ha accesso diretto, e ciò gli consente di conoscerlo con «*near certainty*»; ai processi che plasmano il comportamento, invece, non si ha alcun accesso.²⁵ Poiché i due studiosi non formulano ipotesi circa questa presunta conoscenza diretta di sé, la loro teoria è incompleta.

Una teoria del senso interno

Al fine di offrire una teoria di quella parte della conoscenza di sé che non ha natura interpretativa, alcuni studiosi hanno compiuto, con maggiore o minore radicalità, un ritorno a posizioni più tradizionali, intendendo l'introspezione come un processo che consente di accedere in modo relativamente diretto e non interpretativo ad almeno alcuni eventi mentali.

Secondo queste teorie del senso interno, l'introspezione non consiste nella consultazione di un repertorio teorico al fine di interpretare informazioni extra-mentali; è piuttosto un'operazione che si avvale di meccanismi che ricevono informazioni concernenti il mondo interiore attraverso un canale relativamente diretto.

Il tentativo di conferire plausibilità psicologica alla teoria del senso interno si declina in varie forme. Una versione funzionalista e rappresentazionalista è stata proposta da Shaun Nichols e Stephen Stich.²⁶ I due studiosi distinguono fra il *riconoscimento* (la semplice attribuzione) di uno stato mentale e il *ragionamento* che su tale stato si può condurre, sottolineando poi come entrambe queste operazioni possano essere effettuate in prima e in terza persona. Ora, mentre il riconoscimento in terza persona e il ragionamento in prima e terza persona poggiano sulla medesima ToM, i meccanismi deputati a rilevare i propri stati mentali sono del tutto autonomi dai meccanismi che si occupano degli stati mentali altrui.

Più precisamente, Nichols e Stich ipotizzano l'esistenza di due (o più) *Self-Monitoring Mechanisms* (SMM), uno (almeno) per monitorare e fornire conoscenza dei propri stati percettivi, e un altro per monitorare e fornire conoscenza dei propri stati di atteggiamento proposizionale

(d'ora in poi semplicemente "pensieri").

L'ipotesi degli SMM riguarda esclusivamente l'autoattribuzione di stati mentali. Per quanto riguarda l'eteroattribuzione di stati mentali e il *ragionamento mentalistico* sia in prima che in terza persona, Nichols e Stich riconoscono la validità della TT. Ciò consente loro di delimitare la portata degli esperimenti che attestano effetti di confabulazione: gli errori compiuti dai partecipanti a questi esperimenti riguardano non già l'autoattribuzione di stati mentali, ma il ragionamento mentalistico in prima persona.

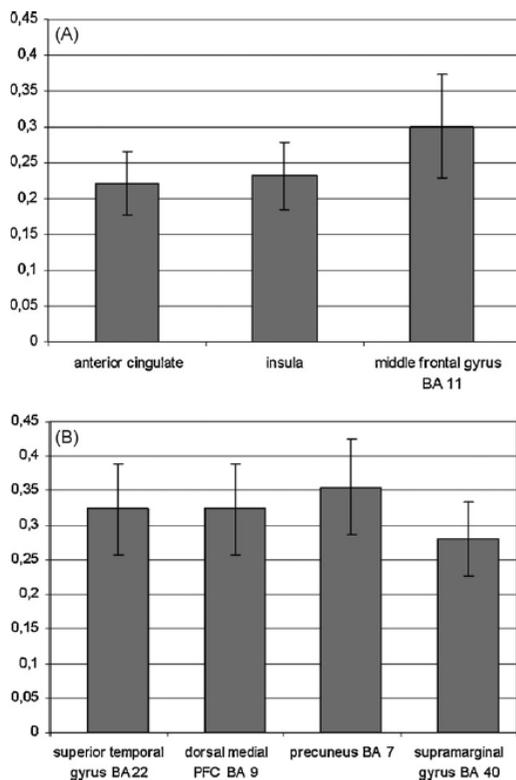


Figura 2. Differenze fra gli schemi di attivazione di pazienti schizofrenici e di controlli normali durante lo svolgimento di compiti di teoria della mente: (A) aree che si sono attivate nei controlli più che negli schizofrenici; (B) aree che si sono attivate negli schizofrenici più che nei controlli. Fonte: M. Brüne, S. Lissek, N. Fuchs, H. Witthaus, S. Peters, V. Nicolas, G. Juckel, M. Tegenthoff, An fMRI Study of Theory of Mind in Schizophrenic Patients With "Passivity" Symptoms, in: «Neuropsychologia», vol. 46, 2008, pp. 1992-2001.

Ossia: comprendere le *cause* del proprio comportamento richiede di ragionare sugli stati mentali, e questo è senz'altro un processo carico di teoria.

Pertanto, se questa teoria (ToM) non ha le risorse necessarie per dar conto di una determinata condotta, il soggetto commetterà errori inferenziali in relazione tanto agli stati mentali altrui che a quelli propri. In altre parole, la conoscenza di sé può contare su *due metodi*: è vero, in talune circostanze il soggetto interpreta in base a ToM (il che può dar luogo a discorsi confabulatori); ma è altresì vero che in altre occasioni egli gode di un accesso alla propria mente che è diretto e non interpretativo.²⁷

Dal momento che la teoria degli SMM asserisce che l'autoattribuzione mentalistica non poggia su ToM, Nichols e Stich prevedono la dissociabilità fra il *mindreading* in prima e in terza persona. Sia diacronica: in luogo del parallelismo previsto dalla teoria della parità io/altro, si dovrebbero rilevare asincronie evolutive.²⁸ Sia sincronica: nei pazienti ASD, gli SMM – e dunque la capacità di riconoscere i propri stati percettivi e i propri pensieri – potrebbero essere intatti a dispetto del deficit che interessa ToM; mentre l'inverso si osserverebbe in pazienti con esperienze di passività – in questo caso la compromissione interesserebbe gli SMM ma non i meccanismi guidati da ToM.²⁹

A queste congetture si può però opporre la considerazione che i dati sopra riportati in favore della teoria della parità io/altro rendono estremamente problematica tanto l'ipotesi di una dissociabilità diacronica fra *mindreading* in prima e in terza persona, che l'ipotesi di una doppia dissociazione fra autismo e schizofrenia. Tanto più che ai dati che attestano anomalie a carico dell'introspezione in pazienti ASD, si somma la letteratura che documenta una stretta associazione fra schizofrenia e deficit del *mindreading* in terza persona.³⁰

In un recente studio fMRI, per esempio, pazienti con esperienze di passività sono riusciti a eseguire semplici compiti di *mindreading* in terza persona, ma ciò è stato realizzato solo mobilitando una rete di aree cerebrali diversa da quella utilizzata dai controlli (fig. 2). Questi

pazienti non possiedono un sistema di mindreading normale, anche se poi riescono in vari modi a compensare (parzialmente) l'anomalia.

■ Il modello dell'accesso sensoriale e interpretativo

Alle critiche or ora esposte nei riguardi della teoria di Nichols e Stich si deve aggiungere quella di Peter Carruthers. Questo studioso ha sostenuto che, se pure l'accesso agli eventi percettivi è relativamente diretto e immediato, l'autoattribuzione dei pensieri ha invece sempre carattere interpretativo, in accordo con quanto sostenuto dalla teoria della parità io/altro. Il risultato è un modello che definisce la conoscenza di sé in termini tanto di accesso sensoriale che di interpretazione: l'*Interpretive Sensory-Access (ISA) Theory of Self-Knowledge*.³¹

Per dar conto dell'accesso agli eventi percettivi il modello ISA si struttura intorno alla teoria dello spazio di lavoro globale neuronale di Bernard Baars. Questa teoria postula una serie di sistemi percettivi che rendono disponibili i loro output (dati sensoriali provenienti dall'ambiente, dati relativi a eventi interni come le immagini mentali e il monologo interiore, dati somatosensoriali e propriocettivi) a una serie di sistemi concettuali "consumatori".

Fra questi compaiono sistemi che utilizzano l'output percettivo per formulare giudizi o per prendere decisioni; e fra i sistemi che generano giudizi vi è il sistema (multicomponenziale) deputato alla lettura della mente che, guidato dal quadro teorico della psicologia ingenua, genera credenze di ordine superiore sugli stati mentali altrui e propri. Il sistema di *mindreading* riceve in input gli stati percettivi trasmessi globalmente, ed è dunque in grado di riconoscere questi percetti e di produrre autoattribuzioni della forma "vedo qualcosa di rosso", "ho sete", "avverto dei brividi", e così via.

Si noti che la disponibilità dei dati percettivi al sistema di *mindreading* è ciò che è legittimo attendersi alla luce della tesi secondo cui il *mindreading* si è originariamente evoluto come capacità di leggere le menti altrui in modo da garantire un vantaggio adattivo nel perseguimen-

to di finalità sociali: anticipare (e talvolta manipolare) il comportamento di conspecifici; migliorare il coordinamento di attività cooperative. Banalmente, i meccanismi di *mindreading* per poter interpretare le azioni altrui devono poter accedere alle loro rappresentazioni percettive.

Il sistema di *mindreading* può dunque accedere a una gran massa di informazioni percettive ma non può attribuirsi direttamente i pensieri, che sono gli output di altri sistemi concettuali: gli eventi di atteggiamento proposizionale *non* sono infatti trasmessi globalmente. Il sistema di *mindreading* deve perciò utilizzare i dati percettivi di cui dispone per *inferire* quali sono i pensieri dell'agente, esattamente come avviene nel caso della lettura delle menti altrui.

Di qui una delle previsioni fondamentali del modello ISA: quando un agente si attribuisce un pensiero ciò avviene *sempre* in virtù di un'inferenza interpretativa istantanea e inconscia, fondata sulla consapevolezza sensoriale di dati concernenti il proprio comportamento, dati contestuali e/o elementi sensoriali contenuti nella memoria di lavoro (per es. un'immagine visiva o una frase nel linguaggio interno).

Carruthers offre un gran numero di argomenti in favore della teoria ISA – fra questi: (i) dati psicopatologici che confutano l'ipotesi di una doppia dissociazione fra auto- ed etero-attribuzione mentalistica (fra cui il lavoro sopracitato di Williams e Happé sull'attribuzione di intenzioni); (ii) dati tratti dalla letteratura sulla confabulazione che estendono la parità io/altro ai pensieri; (iii) considerazioni relative alla funzione evuzionistica del *mindreading* e la letteratura sulla metacognizione.

Si noti che la teoria ISA *predice* la confabulazione dei pensieri; e questa è la sua previsione più importante giacché la distingue sotto il profilo empirico dalle teorie del senso interno. Queste ultime, infatti, possono anche spiegare la confabulazione dei pensieri, ma solo ipotizzando (come fanno Nichols e Stich) un *duplice* accesso ad essi, introspettivo e anche interpretativo; e ciò rende più involuto il loro apparato concettuale.

La teoria ISA, invece, assumendo che il nostro unico accesso ai pensieri passa attraverso

un processo interpretativo guidato da teoria e alimentato da dati sensoriali, comportamentali e contestuali, prevede effetti di confabulazione ogniquale volta i dati sono fuorvianti o la teoria è esplicitamente inadeguata.³²

Come esempio consideriamo lo studio di Joaquim Brasil-Neto e collaboratori.³³ Ai partecipanti veniva chiesto di decidere se muovere il dito indice destro o sinistro allorché udivano un clic. A loro insaputa, il clic era provocato dall'accensione del magnete di un apparato di stimolazione magnetica transcranica, utilizzato per stimolare la corteccia motoria dell'emisfero destro o sinistro, che in tal modo causava il movimento del dito. Sebbene il movimento fosse indotto direttamente dalla stimolazione magnetica, e dunque aggirando il processo decisionale, i soggetti hanno rivendicato la paternità dell'atto, dichiarando di aver consapevolmente *deciso* di muovere il dito.

Si noti: nell'esperimento appena esaminato non è stata presa alcuna decisione; è stato invece causato un comportamento che ha indotto il partecipante a credere di aver preso una decisione. Un teorico del senso interno potrebbe però obiettare: non si può introspeettare quello che non c'è! È solo nei casi in cui *non c'è* una decisione (e di qui generalizzando: in cui non si è formato alcun pensiero) che il soggetto interpreta se stesso; invece, ogni volta che una decisione c'è, è possibile prenderla a oggetto di introspezione. Si tratta però di un'obiezione fragile.

Supponiamo pure che vi siano due metodi distinti con cui attribuiamo decisioni a noi stessi (uno introspeettivo e l'altro interpretativo). Ma non è strano che il metodo interpretativo debba sempre prevalere nei casi in cui non viene presa alcuna decisione? Un meccanismo introspeettivo non dovrebbe avere difficoltà a rilevare un'assenza; e se tale meccanismo segnala che non è stata presa alcuna decisione, mentre nel contempo il sistema di *mindreading* dà luogo all'autoattribuzione di una decisione, non si capisce perché dovrebbe essere il secondo ad avere sempre la meglio sul primo, dando luogo a risposte confabulate alle domande dello sperimentatore.

Insomma, l'introspezione di per sé non ha

alcuna proprietà che possa far ipotizzare che la macchina interpretativa prenda il sopravvento tutte le volte che il meccanismo introspeettivo segnala che nessuna decisione è stata presa. Una simile ipotesi potrebbe aggiungersi alla teoria che postula l'impiego di due metodi per autoattribuirsi le decisioni solo come un assunto ausiliare puramente *ad hoc*.

In favore della teoria ISA Carruthers porta anche alcune considerazioni relative alla filogenesi del *mindreading* e della metacognizione. Una virtù della teoria ISA risiede nella sua parsimonia sul piano della spiegazione evolutivistica: essa postula infatti un unico percorso filogenetico per il *mindreading* sia in prima che in terza persona. Se, come vuole il modello ISA, l'introspezione si struttura nell'atto di volgere su se stessi la capacità di *mindreading* in terza persona, la comparsa della prima capacità costituirà un sottoprodotto dell'evoluzione della seconda.³⁴

Per le teorie del senso interno vi è invece un onere esplicativo. Dal momento che postulano meccanismi neurocognitivi deputati al *mindreading* in terza persona e altrettanti dedicati all'introspezione, queste teorie sono tenute a fornire una spiegazione evolutivistica per la comparsa di ciascuno di tali meccanismi.

Goldman, per esempio, ha proposto che l'introspezione sarebbe stata selezionata per prima; solo in un secondo momento la sua integrazione con processi immaginativi e inferenziali a carattere simulativo avrebbe dato luogo a forme di intelligenza sociale. Quali pressioni evolutive potrebbero aver portato all'avvento della capacità introspeettiva?

Una risposta (probabilmente l'unica che fino a oggi è stata proposta) è che l'introspezione si è evoluta per adempiere a un ruolo di *supervisione* nei riguardi di processi cognitivi di primo ordine: intervenendo in quei processi in caso di difficoltà, avviando nuove strategie, verificando che i compiti venissero svolti secondo le aspettative, e così via. Insomma, essa si sarebbe evoluta per svolgere quell'insieme di compiti cognitivi che è consueto definire "metacognizione".³⁵ Quanto è plausibile questa tesi? Assai poco.

Stando al modello standard dei processi metacognitivi, un metalivello monitora, rappre-

senta e controlla i processi dei sistemi cognitivi del livello oggetto.³⁶ Il metalivello conterrebbe allora un *metamodello* del livello oggetto, ossia costruirebbe *metarappresentazioni* dei processi e degli eventi del livello oggetto. La difficoltà qui è che molti processi metacognitivi *non* hanno carattere metarappresentazionale.

Un esempio è l'impiego di modelli anticipatori nel controllo dell'azione. Secondo l'ipotesi del "sistema comparatore", quando un comando motorio è inviato dal centro alla periferia per dar luogo a un'azione, viene creata una copia efferente delle istruzioni motorie, da cui si origina un modello che predice le conseguenze sensoriali del movimento (ossia una rappresentazione dell'esito previsto). Questo modello predittivo è inviato a un meccanismo, detto per l'appunto "sistema comparatore", che riceve feedback sensoriali dall'azione in corso di svolgimento. Quando il sistema identifica discrepanze fra l'esito previsto e quello che di fatto si sta delineando, attiva un algoritmo che consente di apportare rapide correzioni alle istruzioni motorie (e, conseguentemente, alle contrazioni dei muscoli in periferia) in modo da ottenere una corrispondenza più stretta.³⁷

Ora, il tipo di monitoraggio dell'azione garantito dal sistema comparatore è del *primo ordine*: non richiede la rappresentazione dell'intenzione motoria, e neppure delle esperienze derivanti dall'azione. E ancor meno richiede un accesso introspettivo a intenzioni o esperienze; tanto più che la velocità con cui opera il processo di monitoraggio lascia ben poco margine per l'introspezione.

Dunque, quando i processi metacognitivi non hanno carattere metarappresentazionale, il loro ruolo di supervisione non richiede una capacità introspettiva distinta dal *mindreading* in terza persona. Vi sono però processi che, senz'altro, sono metacognitivi nel senso di essere metarappresentazionali.³⁸ E tuttavia, anche in questo caso si può dubitare che l'introspezione si sia evoluta per svolgere i processi metacognitivi in questione.

A ben guardare, infatti, questi processi sono sì metarappresentazionali, ma è anche vero che operano senza poter intervenire direttamente

sugli stati e i processi rappresentati. Consideriamo le capacità di metamemoria: nella maggior parte dei casi tali capacità si limitano ad avviare un comportamento oppure a intervenire su di esso.

Per esempio, un bambino può preferire un compito di memorizzazione a un altro in base al fatto che il primo contiene un numero minore di elementi: ciò richiede una conoscenza della memoria ma non la capacità di intervenire sul processo mnestico stesso. In modo simile, un soggetto che deve memorizzare alcuni numeri può "ripassarli" nel linguaggio interno: in tal modo egli non interviene direttamente sulla memoria; piuttosto esercita su di essa un'influenza comportamentale indiretta. Tuttavia, non è questo che ci aspetteremmo se la metamemoria fosse una capacità introspettiva che si è evoluta ai fini del controllo esecutivo, consentendo ai soggetti di intervenire direttamente sui processi mnestici.

In conclusione, possiamo dire che l'ipotesi secondo la quale la capacità introspettiva si sarebbe evoluta al fine di svolgere compiti metacognitivi non sembra confermata dai dati. Un esame della letteratura sulla metacognizione rivela che in molti casi il suo ruolo di supervisione non è tale da richiedere una capacità introspettiva separata dalla capacità di lettura delle menti altrui. E rivela altresì che gli interventi metacognitivi non esercitano quell'influenza diretta sull'elaborazione cognitiva che la metacognizione dovrebbe invece esercitare se davvero si fosse evoluta a tale scopo.

Alcune indicazioni per la neuropsichiatria cognitiva

Quanto sinora esposto dovrebbe essere attentamente meditato da tutti quegli psichiatri e psicologi clinici che lavorano nella prospettiva di una *neuropsichiatria cognitiva*: il progetto, vale a dire, di andare oltre i segni e i sintomi della fenomenologia clinica, per sviluppare modelli del disturbo psichico saldamente radicati nel quadro delle scienze neuro cognitive.³⁹

Innanzitutto, uno dei chiarimenti più importanti emersi dalla discussione sopra condot-

ta è la necessità di tracciare una netta linea di demarcazione fra il costrutto della metacognizione e il concetto di introspezione. Al contrario, alcuni approcci al cosiddetto “*mentalization-based treatment*” hanno utilizzato queste nozioni quasi fossero sinonime. In almeno un caso ciò è accaduto saldando l’approccio alla psicopatologia basato sulla mentalizzazione con la sopra discussa teoria del senso interno di Nichols e Stich.⁴⁰

Questi due studiosi assumono infatti l’equivalenza fra metacognizione e introspezione (= automonitoraggio): ciò è evidente allorché citano come evidenza in favore dell’ipotesi dei meccanismi di automonitoraggio una serie di esperimenti sulla metamemoria in bambini ASD (i controlli erano bambini con sviluppo tipico e con ritardo mentale).⁴¹ In base alla teoria della parità io/altro – così ragionano i due studiosi – i bambini autistici avrebbero dovuto trovare maggiore difficoltà dei controlli tanto nei compiti di FB che nei test di metamemoria.

Poiché, al contrario, non si è riscontrata alcuna differenza significativa fra la prestazione metacognitiva degli ASD e quella dei controlli, lo studio parrebbe portare acqua al mulino della teoria dei meccanismi di automonitoraggio. Così non è, tuttavia, e per almeno due ragioni.

Innanzitutto, quasi tutti gli ASD che hanno preso parte alla ricerca sono stati in grado di eseguire compiti di FB del primo ordine; pertanto nulla impedisce al teorico della parità io/altro di prevedere che questi pazienti abbiano conservato una capacità residua di lettura della propria mente, e con essa la capacità di superare semplici test di metamemoria.

In secondo luogo, e fatto ancor più rilevante, nessuno dei test di metamemoria richiedeva al partecipante di attribuirsi pensieri; al contrario, «i compiti potevano essere svolti da chiunque fosse stato in possesso dei concetti mentali necessari e che fosse anche uno scaltro comportamentista».⁴²

Per esempio, un test era volto a stabilire se gli ASD si rendevano conto che fra due insiemi di figure quello contenente il minor numero di elementi è il più semplice da apprendere. Nessuno ha avuto difficoltà a superare questo test,

e ciò si spiega agevolmente alla luce del fatto che in questo esperimento l’età media degli ASD era di 11 anni e quindi avevano trascorso a scuola un periodo di tempo sufficiente per poter stabilire una correlazione attendibile fra il numero di item studiati in un compito e il numero di risposte che sono poi giudicate corrette.⁴³

Dunque, ancora una volta, i dati reclutati a sostegno della teoria dei meccanismi di automonitoraggio si rivelano non problematici per l’*Interpretive Sensory-Access Theory of Self-Knowledge* di Carruthers. E questa teoria parla contro l’ipotesi, avanzata da varie teorie dell’intervento clinico *mentalization-based*, secondo cui il soggetto/paziente si può avvalere di una capacità introspettivo-metacognitiva al fine di accedere ai propri pensieri e di lì arrivare ad esercitare su di essi un controllo *diretto* (per es. regolarli e modificarli attraverso una funzione talvolta denominata “*mastery*”).⁴⁴

Ciò – si noti – è intrinseco a una teoria cognitiva del senso interno: essa infatti *prevede* che il soggetto possieda la capacità di controllare e dirigere quegli stati e quei processi che sono rilevati dal senso interno.

Al contrario, la teoria interpretativista dell’introspezione di Carruthers attribuisce sì alla metacognizione un ruolo di supervisione, ma senza che ciò coinvolga un meccanismo introspettivo distinto dal sistema di lettura della mente altrui. In luogo di un monitoraggio metacognitivo diretto, il modello ISA definisce la metacognizione come un uso in prima persona del sistema di mindreading, il quale si basa per lo più su *cues* suscettibili di accesso sensoriale.⁴⁵ In quest’ottica, alcune “disfunzioni metacognitive” teorizzate dai sostenitori del *mentalization-based treatment* potrebbero rivelarsi degli *artefatti*: infatti, non esiste alcun meccanismo (o insieme di meccanismi) distinto dal sistema di mindreading la cui compromissione sarebbe all’origine di deficit a carattere introspettivo-metacognitivo correlabili con varie tipologie di disturbo mentale (per es. i vari tipi di disturbi di personalità).⁴⁶

Infine, vogliamo concludere segnalando un errore ancora più grave che emerge dalla fiorente letteratura che – assai lodevolmente –

cerca di far interagire *Theory of Mind* e psicopatologia: in taluni casi la metacognizione ingloba tutto il *mindreading*.⁴⁷ Ci auguriamo che il nostro articolo abbia convinto il lettore che, al contrario, i concetti di *mindreading*, introspezione e metacognizione vanno tenuti ben distinti fra loro.

Note

¹ J. PERNER, *Understanding the Representational Mind*, MIT Press, Cambridge (MA) 1991.

² A.M. LESLIE, "Theory of Mind" as a Mechanism of Selective Attention, in: M. GAZZANIGA (ed.), *The New Cognitive Neurosciences*, MIT Press, Cambridge (MA) 2000, pp. 1235-1247.

³ Cfr. A.M. LESLIE, L. THAISS, *Domain Specificity in Conceptual Development: Neuropsychological Evidence from Autism*, in: «Cognition», vol. XLIII, n. 3, 1992, pp. 225-251.

⁴ A sostegno di questa interpretazione si può citare anche uno studio fMRI: cfr. R. SAXE, N. KANWISHER, *People Thinking about Thinking People: The Role of the Temporo-parietal Junction in "Theory of Mind"*, in: «NeuroImage», vol. XIX, n. 4, 2003, pp. 1835-1842. Nella giunzione temporo-parietale destra si è evidenziato un segnale BOLD elevato quando i partecipanti pensavano a credenze erranee; e però l'attività della regione era indistinguibile dai livelli a riposo quando i partecipanti pensavano a foto datate, mappe o indicazioni stradali erranee. Questa potrebbe essere la base neuronale per la dissociazione comportamentale tra la capacità metarappresentazionale mentalistica e quella relativa alle rappresentazioni iconiche.

⁵ Cfr. R. BAILLARGEON, R.M. SCOTT, Z. HE, *False-belief Understanding in Infants*, in: «Trends in Cognitive Sciences», vol. XIV, n. 3, 2010, pp. 110-118.

⁶ Cfr. A.M. LESLIE, *Developmental Parallels in Understanding Minds and Bodies*, in: «Trends in Cognitive Sciences», vol. IX, n. 10, 2005, pp. 459-462.

⁷ Cfr. D. SPERBER, F. CLÉMENT, C. HEINTZ, O. MASCARO, H. MERCIER, G. ORIGGI, D. WILSON, *Epistemic Vigilance*, in: «Mind and Language», vol. XXV, n. 4, 2010, pp. 359-393.

⁸ P. CARRUTHERS, *Mindreading in Infancy*, in: «Mind and Language», vol. XVIII, n. 2, 2013, pp. 141-172.

⁹ Si veda in merito la una descrizione dei tre meccanismi (*Eye Direction Detector; Intentionality Detector; Shared Attention Mechanism*) in S. BARON-

COHEN, *Mindblindness: An Essay on Autism and Theory of Mind*, MIT Press, Cambridge (MA) 1995.

¹⁰ Cfr. A. GOLDMAN, *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*, Oxford UP, Oxford 2006, cap. 7

¹¹ B.J. SCHOLL, A.M. LESLIE, *Modularity, Development and "Theory of Mind"*, in: «Mind & Language», vol. XIV, n. 1, 1999, pp. 131-153, qui p. 140.

¹² Cfr. G. JERVIS, *Psicologia dinamica*, Il Mulino, Bologna 2001, pp. 84-85; G. JERVIS, *Individualismo e cooperazione*, Laterza, Roma-Bari 2002, pp. 173-177. Qui l'autore offre una riformulazione dell'ipotesi dei cinque sistemi motivazionali proposta da J.D. LICHTENBERG, *Psychoanalysis and Motivation*, Analytic Press, Hillsdale (NJ) 1989.

¹³ Cfr. R. BYRNE, A. WHITEN (eds.), *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*, Oxford UP, Oxford 1988; A. WHITEN, R.W. BYRNE (eds.), *Machiavellian Intelligence. Vol. II: Evaluations and Extensions*, Cambridge UP, Cambridge 1997.

¹⁴ P. RICHERSON, R. BOYD, *Not By Genes Alone*, University of Chicago Press, Chicago 2005; S. HRDY, *Mothers and Others*, Harvard UP, Cambridge (MA) 2009.

¹⁵ M. TOMASELLO, M. CARPENTER, J. CALL, T. BEHNE, H. MOLL, *Understanding and Sharing Intentions: The Origins of Cultural Cognition*, in: «Behavioral and Brain Sciences», vol. XXVIII, n. 5, 2005, pp. 675-691.

¹⁶ Cfr. R.E. NISBETT, T.D. WILSON, *Telling More Than we Can Know: Verbal Reports on Mental Processes*, in: «Psychological Review», vol. LXXXIV, n. 3, 1977, pp. 231-259; R.E. NISBETT, L. ROSS, *Human Inference: Strategies and Shortcomings of Social Judgment*, Prentice-Hall, Englewood Cliffs 1980; T.D. WILSON, *Strangers to Ourselves. Discovering the Adaptive Unconscious*, Harvard UP, Cambridge (MA) 2002.

¹⁷ P. JOHANSSON, L. HALL, S. SIKSTRÖM, A. OLSSON, *Failure to Detect Mismatches between Intention and Outcome in a Simple Decision Task*, in: «Science», vol. CCCX, n. 5745, 2005, pp. 116-119.

¹⁸ P. JOHANSSON, L. HALL, S. SIKSTRÖM, B. TÄRNING, A. LIND, *How Something Can Be Said About Telling More Than We Can Know: On Choice Blindness and Introspection*, in: «Consciousness and Cognition», vol. XV, n. 4, 2006, pp. 673-692. Questi risultati sono stati replicati utilizzando materiali del tutto differenti da L. HALL, P. JOHANSSON, B. TÄRNING, S. SIKSTRÖM, T. DEUTGEN, *Magic at the Marketplace: Choice Blindness*

for the Taste of Jam and the Smell of Tea, in: «Cognition», vol. CXVII, n. 1, 2010, pp. 54-61.

¹⁹ E dunque non limitata ai meccanismi periferici di input e output (come, per es., i sistemi deputati alla percezione visiva o al controllo motorio).

²⁰ E. SCHWITZGEBEL, *Introspection*, in: E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, URL = <<http://plato.stanford.edu/archives/win2012/entries/introspection/>>, par. 2.1.3.

²¹ A. GOPNIK, *How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality*, in: «Behavioral and Brain Sciences», vol. XVI, n. 1, 1993, pp. 1-14.

²² H. WELLMAN, D. CROSS, J. WATSON, *Meta-analysis of Theory-of-Mind Development: The Truth About False Belief*, in: «Child Development», vol. LXXII, n. 3, 2001, pp. 655-684

²³ Cfr., tra gli altri, N. FISHER, F. HAPPE', J. DUNN, *The Relationship Between Vocabulary, Grammar, and False Belief Task Performance in Children With Autistic Spectrum Disorders and Children With Moderate Learning Difficulties*, in: «Journal of Child Psychology and Psychiatry», vol. XLVI, n. 4, 2005, pp. 409-419.

²⁴ D. WILLIAMS, F. HAPPE', *Representing Intentions in Self and Other: Studies of Autism and Typical Development*, in: «Developmental Science», vol. XIII, n. 2, 2010, pp. 307-319.

²⁵ R.E. NISBETT, T.D. WILSON, *Telling More Than We Can Know*, cit., p. 255. Questo punto è ribadito in T.D. WILSON, *Strangers to Ourselves*, cit., pp. 17-18. Ma vedi NISBETT e ROSS (Human Inference, cit., pp. 200-202), i quali sembrano escludere la possibilità di accedere introspektivamente tanto alle cause degli stati interni che agli stati stessi.

²⁶ S. NICHOLS, S. STICH, *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, Oxford UP, Oxford 2003, cap. 4.

²⁷ Cfr. A. GOLDMAN, *Simulating Minds*, cit., p. 232.

²⁸ Cfr. S. NICHOLS, S. STICH, *Mindreading*, cit., §§ 4.4.

²⁹ *Ivi*, §§ 4.5-4.6.

³⁰ Cfr. l'accurata rassegna della letteratura di M. BRÜNE, «Theory of Mind» in *Schizophrenia: A Review of the Literature*, in: «Schizophrenia Bulletin», vol. XXXI, n. 1, 2005, pp. 21-42. Nella stessa direzione punta una meta-analisi condotta su 29 studi che hanno coinvolto più di 1500 pazienti: M. SPRONG, P. SCHOTHORST, E. VOS, J. HOX, H. VAN ENGELAND, *Theory of Mind in Schizophrenia: Meta-analysis*, in: «British Journal of Psychiatry», vol. CXCI, 2007, pp. 5-13.

³¹ P. CARRUTHERS, *How We Know Our Own Minds: The Relationship Between Mindreading and Metacognition*, in: «Behavioral and Brain Sciences», vol. XXXII, n. 2, 2009, pp. 121-38; P. CARRUTHERS, *Introspection: Divided and Partly Eliminated*, in: «Philosophy and Phenomenological Research», vol. LXXX, n. 1, 2010, pp. 76-111; P. CARRUTHERS, *The Opacity of Mind: The Cognitive Science of Self-Knowledge*, Oxford UP, Oxford 2011.

³² Cfr. P. CARRUTHERS, *Introspection: Divided and Partly Eliminated*, cit., p. 98.

³³ J. BRASIL-NETO, A. PASCUAL-LEONE, J. VALLS-SOLÉ, L. COHEN E M. HALLETT, *Focal Transcranial Magnetic Stimulation and Response Bias in a Forced Choice Task*, in: «Journal of Neurology, Neurosurgery, and Psychiatry», vol. LV, n. 10, 1992, pp. 964-66.

³⁴ È importante notare, tuttavia, che questo fatto non esclude che in una fase successiva la lettura della propria mente sia stata sottoposta a una selezione secondaria; e si può ipotizzare che ciò sia avvenuto in virtù delle sue funzioni difensive: cfr. T.D. WILSON, *Strangers to Ourselves*, cit.; T.D. WILSON, *Knowing When to Ask: Introspection and the Adaptive Unconscious*, in: «Journal of Consciousness Studies», vol. X, n. 9-10, 2003, pp. 131-140.

³⁵ J. COUCHMAN, M. COUTINHO, M. BERAN, D. SMITH, *Metacognition is Prior*, in: «Behavioral and Brain Sciences», vol. XXXII, n. 2, 2009, p. 142.

³⁶ T. NELSON, L. NARENS, *Metamemory: A Theoretical Framework and New Findings*, in: G. BOWER (ed.), *The Psychology of Learning and Information*, Academic Press, New York 1990, pp. 125-173.

³⁷ C. FRITH, S.-J. BLAKEMORE, D. WOLPERT, *Explaining the Symptoms of Schizophrenia: Abnormalities in the Awareness of Action*, in: «Brain Research Reviews», vol. XXXI, n. 2-3, 2000, pp. 357-363; C. FRITH, *Explaining Delusions of Control: The Comparator Model 20 Years On*, in: «Consciousness and Cognition», vol. XXI, n. 1, 2012, pp. 52-54.

³⁸ Cfr. tra gli altri J. METCALFE, A. SHIMAMURA (eds.), *Metacognition: Knowing about Knowing*, MIT Press, Cambridge (MA) 1994.

³⁹ Cfr. tra gli altri P.W. HALLIGAN, A.S. DAVID, *Cognitive Neuropsychiatry: Towards a Scientific Psychopathology*, in: «Nature Reviews - Neuroscience», vol. II, n. 3, 2001, pp. 209-215; M. MARRAFFA, C. MEINI, *Verso una psichiatria cognitiva*, in: «Sistemi Intelligenti», vol. XVI, n. 2, 2004, pp. 159-183.

⁴⁰ In passato ciò è stato sostenuto in M. FALCONE, M. MARRAFFA, A. CARCIONE, *Metarappresenta-*

zione e psicopatologia, in: G. DIMAGGIO, A. SEMERARI (a cura di), *I disturbi di personalità. Modelli e trattamento*, Laterza, Roma-Bari 2003, pp. 43-76.

⁴¹ A. FARRANT, J. BOUCHER, M. BLADES, *Metamemory in Children With Autism*, in: «Child Development», vol. LXX, n. 1, 1999, pp. 107-131.

⁴² P. CARRUTHERS, *The Opacity of Mind*, cit., p. 305.

⁴³ *Ivi*, pp. 304-305.

⁴⁴ G. DIMAGGIO, A. SEMERARI, *I disturbi di personalità*, cit., pp. 50.

⁴⁵ La natura inferenziale e basata su *cues* sensoriali dei giudizi metacognitivi è un dato consolidato nella letteratura sulla meta cognizione; cfr. tra gli altri J. DUNLOSKY, J. METCALFE, *Metacognition*, Sage, Thousand Oaks (CA) 2009.

⁴⁶ Cfr. per esempio G. DIMAGGIO, A. SEMERARI, *I disturbi di personalità*, cit., parte II.

⁴⁷ Cfr. il recente G. DIMAGGIO, P.H. LYSAKER (a cura di), *Metacognizione e psicopatologia. Valutazione e trattamento*, Raffaello Cortina, Milano 2011.