



UNIVERSITA' DEGLI STUDI DI MESSINA

Dottorato di ricerca in Scienze Cognitive

ciclo XXIX

Il ruolo della flessibilità cognitiva nella costruzione delle abilità di teoria della
mente

Dottoranda:

Francesca Giuliano

Coordinatore:

Ch.mo Prof. Antonino PENNISI

Relatore:

Ch.mo Prof. Marco MAZZONE

Anno Accademico 2015 - 2016

Indice

Introduzione.....	1
-------------------	---

I. Capitolo. Teoria della mente ed *early mindreading*

1.1 Teoria della mente e primati non umani.....	5
1.2 Teoria della mente nell'uomo.....	12
1.3 In che modo è possibile valutare la presenza di abilità di teoria della mente.....	15
1.4 Discussione delle posizioni sull' <i>early mindreading</i>	19

II. Capitolo. Flessibilità cognitiva, apprendimento associativo e *mindreading*

2.1 Flessibilità cognitiva e modularità massiva.....	37
2.2 Flessibilità cognitiva e funzioni esecutive.....	43
2.3 Flessibilità cognitiva e meccanismi automatici.....	45
2.4 Flessibilità cognitiva e meccanismi gerarchico- associativi.....	50
2.5 Flessibilità cognitiva e <i>mental file</i>	59

2.6 Flessibilità cognitiva, associazioni e generalizzazioni.....	65
2.7 Flessibilità cognitiva e teoria della mente.....	72
2.8 Flessibilità cognitiva, coscienza e modularismo.....	81
2.9 Reti di bayes: è ancora utile la coscienza?.....	82

III. Capitolo. Approfondimento critico

3.1 <i>Mental file</i> e <i>belief file</i> : un confronto tra la posizione di Kovacs e quella di Perner.....	86
3.2 <i>Dual systems</i> e <i>multi-systems</i> : due modelli a confronto.....	99
3.3 <i>Bayesian change-point analysis</i> : i limiti del metodo statistico tradizionale.....	106

Conclusione.....	110
-------------------------	------------

Bibliografia.....	112
--------------------------	------------

Introduzione

La capacità di comprendere ed attribuire stati mentali (intenzioni, credenze, desideri) è definita in letteratura teoria della mente o, più in generale, *mindreading*.

Questa capacità ci permette di comunicare (nel senso più ampio del termine) e di percepire gli altri come agenti intenzionali che possiedono stati mentali che possono differire dai nostri.

Lo scopo di questo lavoro è quello di far luce sulle modalità attraverso le quali questa abilità si manifesta nell'individuo, ponendo l'attenzione sul ruolo che la flessibilità cognitiva assume nella costruzione di tale processo (abilità). Superando il paradigma modularista, per diversi anni dominante in ambito psicologico, sarà possibile pensare ad un modello cognitivo costruito su un architettura di tipo bayesiano, in cui gli stati mentali non saranno altro che generalizzazioni sulla base dei dati provenienti dall'esperienza (Mazzone 2014). La tesi è divisa in tre capitoli. Il primo capitolo sarà dedicato all'analisi della letteratura sul *mindreading*. Verrà ripreso il dibattito sulla teoria della mente a partire dalle sue origini attraverso un confronto tra la teoria della mente nei primati non umani e nell'uomo. Verranno ripercorse le posizioni presenti in letteratura, passando per lo storico dibattito che contrappone sostenitori della *theory of theory* e sostenitori della *simulation theory*, fino ad arrivare al cuore della tesi che coincide con l'interesse attuale della ricerca: lo studio delle precoci abilità di *mindreading* (*early mindreading*). Questo cambiamento dell'oggetto di ricerca prende avvio con un articolo di Onishi & Baillargeon (2005). I due studiosi, attraverso una serie di esperimenti, arrivano a sostenere che i bambini sarebbero in possesso di capacità di metarappresentazione già a partire dai 15 mesi.

L'abbassamento dell'età in cui i bambini possederebbero abilità di teoria della mente (prima di Onishi & Baillargeon 2005 si pensava che queste abilità fossero presenti nei bambini solo a partire dai quattro anni), ha acceso un dibattito volto ad indagare le modalità attraverso le quali è possibile leggere il comportamento e le intenzioni altrui così precocemente.

Le posizioni che verranno analizzate in letteratura saranno divise in due macro gruppi: da una parte gli approcci di matrice essenzialmente innatista (Carruthers 2005; Leslie 2005; Leslie, Polizzi, German, 2005), e dall'altra gli approcci *behaviour-rules based* (Ruffman & Perner 2005; Perner 2010; Perner 2014; Ruffman 2014).

I membri del primo gruppo, ritengono che gli esperimenti di Onishi & Baillargeon rivelino la presenza di early *mindreading* nei bambini. Sostengono dunque che a sancire la differenza tra il precoce e il più tardo *mindreading*, sia la maturazione di abilità, differenti da quelle di mentalizzazione, non presenti nei bambini molto piccoli.

I sostenitori degli approcci *behaviour-rules based* ritengono invece che nei bambini non siano presenti in età precoce abilità di *mindreading* e che gli esperimenti di Onishi & Baillargeon siano spiegabili attraverso la sola presenza di abilità di tipo percettivo-associativo.

All'interno di questi due macro gruppi, in una posizione ibrida tra i primi approcci e i secondi, si situano le posizioni di Apperly & Butterfill e Christensen & Michael. I primi giustificano le precoci abilità di *mindreading* ipotizzando l'esistenza di due sistemi qualitativamente differenti l'uno dall'altro.

Il primo, garantirebbe efficienza e velocità. Il secondo, il più complesso, garantirebbe flessibilità e abilità inferenziali di tipo metarappresentazionale. Nei bambini molto piccoli sarebbe presente solo il primo sistema.

Christiensen & Michael propongono piuttosto una cooperazione tra sistemi che interagiscono tra loro scambiandosi informazioni.

Nel secondo capitolo verranno analizzate le posizioni presenti in letteratura sulla flessibilità cognitiva che oscillano tra chi vede la flessibilità dipendente da strutture innate (i sostenitori della modularità massiva) e posizioni che la vedono dipendente da strutture perlopiù apprese e costruite “storicamente” (i sostenitori del modello gerarchico associativo, che considerano i meccanismi controllati alla base dell’agire flessibile, e i sostenitori dei modelli che fondano la flessibilità su meccanismi automatici).

Verrà prestata particolare attenzione ai modelli gerarchico-associativi, e alle modalità attraverso le quali la letteratura definisce l’idea stessa di associazione. La posizione di Tenenbaum (2011) verrà presa in considerazione soprattutto per la sua argomentazione sul modo in cui da contesti particolari è possibile estrarre principi astratti applicabili a situazioni molteplici che, come vedremo, è funzionale allo scopo ultimo di questo lavoro. Verrà ripresa la posizione di Smith (2000) nel suo tentativo di riabilitare l’associazionismo distinguendolo dal comportamentismo, e verranno presi in considerazione gli studi di Rakison e Lupyan (2008) che, criticando aspramente gli approcci sulla core knowledge, hanno provato a formulare un modello cognitivo, definito *constrained associative learning* (CAAL), fondato sui concetti chiave di astrazione, associazione e immagazzinamento di informazioni. Questa rivisitazione dell’idea di associazione, in un’architettura cognitiva di tipo bayesiano, verrà affiancata all’idea di “concettualizzazione situata” di Barsalou, che lui stesso ha recentemente ha applicato al *social priming*.

Nel terzo capitolo, infine, verranno approfondite alcune delle questioni già prese in considerazione nei precedenti capitoli. Verrà proposto un confronto tra le posizioni di Kovacs (2015) e di Perner (2015) a proposito dell’idea di mental file e in seguito verrà riportato il confronto tra “sostenitori del dual system” (Apperly

& Butterfill) e sostenitori della *multi-systems architecture* (Christiansen & Michael).

Un ultimo paragrafo sarà infine dedicato ad una questione molto attuale che riguarda le metodologie utilizzate per la valutazione dei risultati nei compiti di falsa credenza. Verrà posta l'attenzione su un metodo statistico di tipo bayesiano (*Bayesian change-point analysis*) che fornirà una risposta alternativa alle questioni sollevate nel primo capitolo riguardo all'età in cui le abilità di metarappresentazione si presentano per la prima volta nel bambino.

Capitolo 1

Teoria della mente ed early mindreading

1.1 Teoria della mente e primati non umani

La capacità di prevedere e leggere il comportamento e le intenzioni altrui (*mindreading*) è alla base delle interazioni sociali. Sarebbe impossibile pensare a qualsiasi scambio comunicativo (anche non linguistico) prescindendo dall'utilizzo di strumenti cognitivi adibiti a tali funzioni di tipo predittivo-interpretativo. Da un punto di vista evolucionistico è chiaro come gli organismi in possesso di abilità inferenziali fossero premiati dalla selezione naturale. Tra un organismo in grado di prevedere, ad esempio, le mosse del nemico ed uno privo di tale abilità, è chiaro come il primo fosse enormemente più avvantaggiato. É impossibile negare dunque il valore adattativo di tale capacità che, seppur già posseduta da altre specie animali, nell'uomo si presenta ad un grado superiore di complessità. Premack & Woodruff in un articolo del 1978 *Does the chimpanzee have a theory of mind?* per primi si interessarono alla possibilità di comparare le capacità inferenziali dei primati a quelle umane. Ne conclusero che anche gli scimpanzé possedevano una teoria della mente. Le conclusioni a cui arrivarono i due studiosi vennero facilmente criticate. Lo stesso Premack (1983) ridimensiona le affermazioni fatte in precedenza pur non rinnegandone le conclusioni:

In complesso ci sono solo prove indicative dell'esistenza di una teoria della mente negli scimpanzé (o in qualsiasi non umano). Inoltre, anche le prove positive lasciano aperta la possibilità di una teoria della mente più debole di quella che vale per gli esseri umani. Primo, se mai le scimmie antropomorfe attribuiscono stati mentali questi sono certamente un piccolo sottoinsieme di quelli che vengono attribuiti dagli esseri umani. Secondo, queste scimmie

potrebbero non essere in grado di attribuire stati di conoscenza diversi da quelli propri. Potremmo quindi dover distinguere tre gradi di teoria della mente: (a) specie che non compiono attribuzioni di alcun genere, presumibilmente il caso che si applica a quasi tutte le specie; (b) specie le cui attribuzioni sono illimitate in ogni senso eccetto forse per il numero di connessioni (ad esempio: John pensava che Mary credeva che Bill pensava che...), ipoteticamente il caso degli esseri umani (da quando hanno 4 anni di età) (si veda Wimmer e Perner 1983); e (c) specie che compiono attribuzioni che sono però limitate in molti modi, probabilmente il caso degli scimpanzé.

Tomasello, Call (2008) nel celebre articolo *Does the chimpanzee have a theory of mind? 30 years later* rispondono alla domanda posta diversi anni prima da Premack e Woodruff. Se è vero che gli scimpanzé possiedono un sistema cognitivo tale da permettere loro di leggere il comportamento di altri agenti intenzionali, è anche vero che è improbabile che essi ragionino in termini di falsa credenza:

In a broad construal of the phrase 'theory of mind', then, the answer to Premack and Woodruff's pregnant question of 30 years ago is a definite yes, chimpanzees do have a theory of mind. But chimpanzees probably do not understand others in terms of a fully human-like belief–desire psychology in which they appreciate that others have mental representations of the world that drive their actions even when those do not correspond to reality. And so in a more narrow definition of theory of mind as an understanding of false beliefs, the answer to Premack and Woodruff's question might be no, they do not. Why chimpanzees do not seem to understand false beliefs in particular – or if there might be some situations in which they do understand false beliefs – are topics of ongoing research. (Tomasello & Call, 2008:191)

Una posizione nettamente in contrasto con quella di Premack & Woodruff e con quella di Tomasello & Call è quella di Povinelli & Vonk (2003). Secondo questi ultimi i primati superiori non possederebbero nessuna teoria della mente ma agirebbero in virtù di astrazioni comportamentali. La capacità di mentalizzazione secondo i due studiosi sarebbe una prerogativa della specie umana:

Humans and chimpanzees undoubtedly inherited common mental structures for forming behavioral abstractions. During the course of hominid evolution, however, our lineage might have woven a new, 'theory-of-mind system' into our ancestral cognitive architecture in such a way that the new and old systems now exist in perfect harmony alongside each other – just as the numerous systems sub-serving echolocation in bats were woven in alongside mental systems that they share with their closest relatives. We have labeled this the 'reinterpretation hypothesis' to emphasize that in this scenario, the capacity for behavioral abstraction was already present in the common ancestor, and that humans added another system for coding the behaviors in an additional, mentalistic fashion. (Povinelli & Vonk,2003:2)

Quello che stanno dicendo Povinelli e Vonk è che esiste una differenza qualitativa tra le abilità inferenziali dei primati superiori (possibili attraverso “astrazioni del comportamento”) e quelle umane (possibili attraverso “un nuovo sistema di teoria della mente”). É proprio su questo punto che Tomasello, Call e Hare muovono delle critiche:

Povinelli and Vonk have adopted a kind of derived behaviorism, that is, they note that organisms must observe behavior in order to attribute to it underlying psychological processes, and so it is always possible that the

organism is simply making 'behavioral abstractions'. With regard to one of our food-competition studies in particular, Povinelli and Vonk point out some possible behavioral cues that the subordinate chimpanzee might have used to determine whether the dominant chimpanzee had seen food over which the two of them were competing (and use this information to decide whether to pursue that food). These are such things as: 'avoid food when a specific individual has previously oriented to it in its current location (and not in another location)'. How chimpanzees might have acquired such rules is not specified, nor is there any speculation about how they might understand them. (Tomasello, Call, Hare, 2003: 239)

E ancora continuano:

Of course it is possible that human beings are the only species that understand any psychological processes in others, and we ourselves held this position not so very long ago. But evidence is mounting that it is simply not the case. We cannot dismiss this evidence by noting that simpler explanations are hypothetically possible with no supporting evidence. And, to repeat our earlier, more general point, we are certainly never going to make progress on questions concerning the evolution and ontogeny of social cognition if we think in terms of a monolithic 'theory of mind' that species either do or do not have. We are beyond that now. It is time to turn up the microscope. (Tomasello, Call, Hare, 2003:239-240)

Dire che le abilità inferenziali dei primati si reggono su astrazioni comportamentali, e sostenere che un sistema di teoria della mente è presente solo nella specie umana, significa per Povinelli e Vonk ragionare in nome del rasoio di Occam. Perché chiamare in causa abilità complesse, come quelle di mentalizzazione, se già i meccanismi associativi e le euristiche comportamentali

sono sufficienti a giustificare le abilità dei primati? In un certo senso secondo Povinelli e Vonk, Tomasello e Call risolverebbero in maniera dispendiosa e complessa una questione che potrebbe essere risolta in maniera molto più semplice.

La proposta però, come abbiamo detto, non convince i teorici del *mindreading* che giudicano questa spiegazione oltre che insufficiente, erronea. Dire che le abilità inferenziali dei primati differiscono qualitativamente da quelle dell'uomo, significa dire che esiste un "salto evolutivo" tra le abilità dei primi e quelle dei secondi. In questo modo Povinelli e Vonk offrirebbero un quadro fuorviante del processo di evoluzione naturale. Quest'ultima secondo Tomasello sarebbe fondata sull'idea di continuità e non su quella del "tutto o nulla"¹. In uno dei suoi ultimi lavori Tomasello (2014) si serve di una serie di esperimenti per arrivare a sostenere che le grandi antropomorfe sono capaci di inferenze che vanno oltre la spiegazione di matrice comportamentista. Esse infatti:

sembrano monitorare non solo azioni simulate e i risultati simulati di queste, o cause simulate e gli effetti simulati di queste, ma anche il proprio bagaglio di conoscenza o di memoria, che usano poi per fare inferenze sulle proprie probabilità di successo comportamentale. Le grandi antropomorfe e gli altri primati possono in qualche modo accedere, almeno in contesti strumentali, ai propri stati psicologici. E anche se questa capacità differisce dalla versione compiutamente umana dell'autoriflessione (giacché, come vedremo, manca della dimensione sociale/prospettica), essa mostra che le grandi

¹Con questo non si intende dire che non esistono valide teorie evolucionistiche che contemplano la discontinuità evolutiva. Si vuole rispettare piuttosto il principio seguito dallo stesso Tomasello (2014:29): "Poiché l'uomo e le altre antropomorfe hanno condiviso la storia evolutiva fino a tempi recenti – e condividono ancora, in buona parte, struttura fisica organi di senso, emozioni e organizzazione cerebrale- adatteremo, in mancanza di prove del contrario, un principio di continuità evolutiva"[de Waal 1999].

antropomorfe utilizzano tutte e tre le componenti principali-rappresentazioni cognitive astratte(modelli), paradigmi inferenziali protologici, automonitoraggio e valutazione psicologici- di quello che non può che essere chiamato altrimenti che pensiero. (Tomasello, 2014: 42)

Martin, Santos (2014) sulla stessa linea di Tomasello (2014), chiariscono in maniera più precisa la questione: se è vero che i primati non umani ragionano sul comportamento dei conspecifici lasciando trapelare una certa comprensione degli stati mentali altrui, non c'è nessuna evidenza che confermi che i primati non umani siano capaci di comprendere le credenze altrui.

Ciò che è davvero convincente di questi studi è la tipologia di test che viene utilizzata per arrivare a sostenere questa tesi. I primati vengono infatti sottoposti agli stessi test a cui sono stati sottoposti i bambini in età prescolare, utilizzando uno dei paradigmi più frequenti, quello che si basa sulla fissazione dello sguardo²:

In this experiment, monkeys in the true belief condition saw a human experimenter watch an object move into one of two box hiding locations. Like 15-month-old infants, monkeys in this condition looked longer when the experimenter reached in the incorrect location than when she reached in the correct location. In the false belief condition, monkeys first saw an event in which a human experimenter watched an object move into one of two boxes. The experimenter then turned away as the monkey alone saw the object move from the original box to the second box. In contrast to human infants' performance, monkeys looked for the same duration when the experimenter with a false belief reached to either of the two location. (Martin, Santos,2014: 4)

² Le abilità precoci di *mindreading* verranno trattate approfonditamente nei paragrafi successivi.

In uno studio recentissimo Martin, Santos (2016) trovano ulteriori conferme all'idea che le abilità di *mindreading* espresse dai primati non umani non possano essere paragonate a quelle umane. In particolare da questo studio risulta ancor più evidente che i primati sarebbero capaci di percepire gli altri simili sulla base di percezioni che riguardano lo stato di cose attuale e passato ma non sarebbero, nonostante tutto, capaci di attribuirgli credenze.

Il fallimento nei compiti di falsa credenza (impliciti), indica che i primati sono in grado di comprendere ciò che risulta essere vero esclusivamente dalla loro prospettiva ma non di comprendere (stati mentali altrui che rappresentano) ciò che per loro risulta falso, con buona pace di chi sosteneva il contrario. Queste abilità permetterebbero loro di competere con i loro simili, anticipare le loro azioni in una situazione di caccia o di pericolo, e di compiere tutte quelle azioni necessarie alla loro sopravvivenza. I due studiosi arrivano dunque a concludere:

Primates understand something critical about the relation between agents and information: primates can represent relations between agents and information that is true from their own perspective. Such awareness relations allow primates to functionally exploit what others know and do not know, and to make correct predictions about others' future behavior. Importantly, however, there is also a critical limit to the awareness relations that primates can represent: primates cannot represent relations between agents and untrue or decoupled states of the world. These more computationally sophisticated representational relations allow human ToM to go beyond that of other primates. In doing so, we may have become the only species that is able to track the contents of others' minds even when the contents of those other minds differ from our own. (Martin, Santos,2016:7)

1.2 Teoria della mente nell'uomo

In che modo si sviluppano le abilità di teoria della mente e quando queste ultime si manifestano nell'individuo, è oggetto di studio da diversi anni.

Ripercorrendo diacronicamente l'intera questione relativa alle abilità di *mindreading*, è possibile assistere ad un cambiamento dell'oggetto principale della ricerca.

Se all'inizio si indagavano soprattutto le modalità attraverso le quali era possibile comprendere le intenzioni altrui, nell'ultimo periodo si cerca sempre di più di individuare un momento specifico nello sviluppo del bambino durante il quale tali abilità si manifestano.

Si moltiplicano così ricerche al fine di trovare una discriminante che permetta di distinguere con esattezza le abilità di metarappresentazione *tout-court* dalla percezione di semplici associazioni di tipo percettivo-associativo.

E' interessante notare a tal proposito, che i termini *mindreading* e *theory of mind* vengono ormai utilizzati perlopiù come sinonimi, mentre in passato l'utilizzo del termine *theory of mind* esprimeva l'adesione ad una particolare ipotesi sulla comprensione e sullo sviluppo di queste capacità, quella che presupponeva appunto il possesso di una teoria mentale: la "teoria della teoria" (*theory of theory*). Quest'ultima nello storico dibattito si contrapponeva alle teorie fondate sulla simulazione mentale (*simulation theory*)³. Tra i più ferventi sostenitori della *theory of theory* figurano coloro che credono che le abilità di metarappresentazione siano possibili grazie alla presenza di un modulo innato di psicologia ingenua, i modularisti (Fodor, Leslie, Baron-Cohen). L'architettura cognitiva necessaria a permettere operazioni di questo tipo è specifica per dominio.

³ Un'ulteriore possibilità di ridurre e raggruppare le due posizioni presenti in letteratura è quello di contrapporre innatisti e modularisti (*nativist modular accounts*) da una parte, e costruttivisti (*constructivist conceptual shift accounts*) dall'altra (Low, Perner, 2012)

Un modo differente di intendere la *theory of theory*, è quello di coloro che ritengono che a garantire le abilità di *mindreading* sia un meccanismo generale per dominio e che queste abilità seguano leggi strutturate come le leggi scientifiche (Gopnik, Meltzoff, Ruffmann).

Questa ulteriore ramificazione della “theory of theory” è stata anche definita “teoria del piccolo scienziato” (“the child as scientist theory,”). A caratterizzare questa teoria è il fatto di concepire le abilità di metarappresentazione come il frutto di un processo di apprendimento fondato su un meccanismo di “prova ed errore” (*trial and error*). Esattamente come avviene nelle teorie scientifiche, valide finché non falsificate, il bambino esplorerebbe il mondo sociale sperimentando e costruendo passo dopo passo, errore dopo errore, la sua conoscenza sul quel mondo.

Dunque esattamente come lo scienziato moderno, il bambino davanti ad un problema (o a qualcosa che non conosce) elabora ed utilizza una strategia (una teoria) per risolverlo. Questa teoria costruita attraverso la risoluzione e la correzione di errori, sarà valida fino a quando un'altra non la falsificherà.

La teoria della simulazione (Goldman, Gordon), si fonda invece sull'idea che la percezione dell'altro sia possibile “mettendosi nei suoi panni”. Non sarebbe dunque una teoria, almeno non così come intesa dai teorici della *theory of theory*, a garantire abilità di *mindreading*, ma una forma di simulazione mentale che permetterebbe di interiorizzare gli stati mentali altrui. Questi ultimi sarebbero percepiti in prima persona e non più in terza persona attraverso la sensazione di provare quello che l'altro sta provando.

La celebre scoperta dei neuroni specchio (Rizzolatti, Sinigaglia, 2006) è forse la più felice conquista per i teorici della simulazione mentale. I neuroni specchio sono neuroni che si attivano sia quando il soggetto compie un'azione sia quando questi osserva la sua esecuzione. Ciò testimonierebbe in favore dell'esistenza di un legame tra l'osservazione delle azioni degli altri e la nostra capacità di

interpretarle. Questa scoperta ha fornito preziosi spunti per le successive ricerche.

Un'altra parte delle scienze cognitive attribuisce una grande importanza al corpo e alla narrazione di storie (pratiche narrative) nello sviluppo di abilità di metarappresentazione nel bambino. Le teorie che prendono spunto dalla fenomenologia e dal costruttivismo di Vygotskij, fanno a meno della mentalizzazione (per una rassegna vedi Gallagher, Zahavi, 2008).

L'influenza della tradizione fenomenologica è evidente in queste teorie nel momento in cui viene negata la necessità di ammettere l'esistenza di un meccanismo cognitivo alla base dell'approccio intersoggettivo. Secondo i sostenitori di queste teorie, la comprensione dell'altro non avverrebbe attraverso una forma di simulazione, né grazie al possesso di una teoria. Non vi sarebbe nessuna conoscenza intersoggettiva mediata dalla mentalizzazione, ma essa sarebbe piuttosto immediata, e possibile perché "socialmente contestualizzata" (Gallagher, Zahavi 2008; trad.it:267). Bruner e Hutto, sono i principali teorici delle pratiche narrative, oltre a Gallagher e Zahavi (2008), che hanno il merito di aver ripreso la tradizione fenomenologica.

Bruner (2002), crede che le pratiche narrative costituiscano una dimensione importante della cognizione sociale al punto da fornire una valida alternativa alle troppo dispendiose teorie mentali. I bambini a partire dall'età di due anni agirebbero sulla base dell'attribuzione di narrazioni per spiegarsi il comportamento altrui.

Colui che sviluppa ulteriormente l'idea generale di Bruner, è Daniel Hutto, il quale nella sua Narrative Practice Hypothesis (NPH), illustra come sia determinante la pratica del raccontare storie nello sviluppo della comprensione intersoggettiva da parte del bambino.

Approcci per alcuni versi molto simili, sempre di matrice vygotiskijana, cercano di comprendere il ruolo che il linguaggio posseduto dalla figura di riferimento per il bambino (*caregiver*) possa avere nello sviluppo di tali abilità. Secondo questi approcci, la partecipazione alle interazioni comunicative è la responsabile del successivo sviluppo di abilità di *mindreading* (Nelson 2005; Dunn & Brophy 2005; Harris 2005). In approcci di questo tipo la qualità dell'interazione con la figura di riferimento è una variabile estremamente rilevante e la capacità da parte della madre di comprendere e verbalizzare gli stati mentali del bambino (*mind-mindness*) può divenire lo specchio attraverso il quale egli stesso comprenderà gli altri (Meins,2012).

1.3 In che modo è possibile valutare la presenza di abilità di teoria della mente

La capacità di comprendere credenze è stata storicamente valutata attraverso il famoso test della falsa credenza (Wimmer and Perner 1983).

La scena alla quale il soggetto assiste è la seguente⁴: Maxi, mette il cioccolato in una scatola blu. In sua assenza la madre sposta il cioccolato in una scatola verde. Il bambino deve indicare dove Maxi cercherà il cioccolato al suo ritorno.

Prima dei quattro anni i bambini non riescono a superare il test della falsa credenza e tendono ad indicare la scatola verde. Fanno cioè coincidere la propria credenza con quella di Maxi a causa del fatto che non riescono a riconoscere che la loro prospettiva (che è quella di chi ha assistito alla scena e quindi sa che la madre ha spostato il cioccolato) è diversa da quella di Maxi (che non avendo visto la madre entrare e spostare il cioccolato, lo cercherà dove lui stesso lo ha riposto).

⁴ Sono state sviluppate diverse versioni di questo test. Quest'ultimo è noto anche come il test di Sally e Anne, le due protagoniste della scena (Baron-Cohen et. al.,1985)

Il bambino che non supera questi test non è in grado di tenere a mente due differenti rappresentazioni, la sua e quella dell'altro agente e non riesce a vedere quest'ultimo come un essere pensante e indipendente.

Per diversi anni questi esperimenti ripetuti in molteplici varianti, ad esempio quella in cui viene cambiato il contenuto di un oggetto rispetto a quello che è solito contenere (per esempio quello in cui si sostituisce il contenuto di una scatola di matite con delle candele, o il contenuto di una scatola di caramelle con dei pennarelli), non hanno fatto altro che confermare l'idea che prima dei quattro anni non fosse possibile comprendere stati mentali.

Lo scopo del test (Wimmer, Perner, 1983) stando a quello che ne dicono gli autori, era quello di valutare se i soggetti avessero una rappresentazione della falsa credenza *esplicita* e definita.

É proprio su questo punto che le successive ricerche si sono concentrate: è forse possibile che le abilità di metarappresentazione siano presenti nel bambino ad uno stadio precedente del suo sviluppo? É possibile cioè che esista uno stadio intermedio tra quella consapevolezza dell'interiorità altrui e la sua totale incapacità di percepirla? Che tipo di test devono essere pensati per fare in modo che anche in bambini più piccoli sia possibile valutare se esistono abilità di questo tipo?

Si inizia insomma a pensare a differenti test, più semplici e immediati, che coinvolgano il più possibile ed esclusivamente abilità di metarappresentazione, e che siano capaci di verificare il possesso di abilità di *mindreading* ad uno stadio meno avanzato. Si iniziano così a distinguere le abilità *esplicite* da quelle *implicite*⁵:

⁵ Il dibattito attuale sulla *theory of mind* si concentra principalmente su questa distinzione.

The distinction between implicit and explicit knowledge plays an increasingly important role not only in cognitive development but in cognitive science at large, despite the fact that no agreed meaning of this distinction has yet emerged. Our use is primarily descriptive and intuitive. We illustrate the distinction with examples from children's intellectual development: There are now several studies suggesting that implicit knowledge precedes explicit understanding. One important source of evidence are findings that children learn to perform particular tasks before they can understand and talk about what they are doing. (Perner, Clement,1994:378-379)

Il termine implicito è utilizzato dunque per indicare il possesso di una teoria della mente rudimentale e indefinita mentre il termine esplicito indica il possesso di una abilità ad un grado più avanzato del suo sviluppo e dunque più sofisticata. Alla base di questa distinzione c'è dunque l'idea che comprendere gli altri come agenti intenzionali sia un'abilità che si sviluppa su differenti livelli di complessità:

[...] We do not want to demonstrate earlier understanding of the same kind (i.e., explicit) as measured in the original belief test, but demonstrate understanding at a different level (i.e., implicit). To date all variations of the original false belief test assess explicit understanding in the sense that children are asked a question about the protagonist's belief or action that they have to answer. In this study, we develop a measure of implicit understanding that does not depend on the children's answering a question. Instead, their automatic anticipation of story events is exploited. We simply recorded where children looked in anticipation of the protagonist reappearing. To make this anticipatory looking as clear as possible, the original story needed only minor changes. (Perner, Clement,1994:379)

In questo tipo di indagini si valuta la risposta *spontanea* del bambino di fronte a compiti per la cui riuscita è necessario il possesso di abilità di metarappresentazione.

Le metodologie utilizzate per valutare le abilità implicite sono lo sguardo anticipatorio (*anticipatory looking*) e la violazione dell'aspettativa (*violation of expectation*). I test utilizzati da Perner, Clement (1994) si fondano sullo sguardo anticipatorio.

La situazione a cui i bambini assistono è la seguente: un topo depone il suo formaggio in un luogo x. Mentre il topo non è presente, il formaggio viene spostato in un luogo y e al bambino viene detto che il topo sta tornando e vorrà riprendere il suo formaggio. Ciò che viene osservato è lo sguardo del bambino. Se quest'ultimo soffermerà il suo sguardo nel posto in cui il formaggio realmente si trova, si potrà sostenere che il bambino non possiede abilità di *mindreading*, se il bambino prevede le mosse del topo e quindi guarda nel luogo in cui il formaggio era stato posto inizialmente, cioè dove il topo crede che sia, si potrà sostenere che il bambino è in grado di comprendere la falsa credenza del topo.

E' stato dimostrato nei seguenti esperimenti che solo a partire dai 2 anni e 11 mesi i bambini fissavano maggiormente lo sguardo nella situazione di falsa credenza, a dimostrazione che già a quell'età sono presenti rudimentali capacità di comprensione della mente, mentre tra i 2 anni e 5 mesi e i 2 anni e 10 mesi i bambini soffermavano lo sguardo nell'ultimo luogo in cui avevano visto depositare il formaggio. In base, dunque, alla loro credenza. Sono stati creati altri esperimenti (Clements & Perner, 1994; Clements et al., 2000; Garnham & Perner, 2001; Garnham & Ruffman, 2001) e sono state confermate le ipotesi già avanzate da Perner & Clement (1994).

In seguito altri esperimenti hanno valutato sempre attraverso l'utilizzo di paradigmi simili, se fosse possibile riscontrare abilità di *mindreading* in bambini ancora più piccoli. Così dai due anni e 11 mesi l'età in cui sono state testate abilità di teoria della mente si è abbassata prima ai 2 anni (O'Neill 1996), e poi addirittura ai 18 mesi (Repacholi & Gopnik 1997)⁶. L'articolo che ha fatto sorgere dubbi sulla reale natura e sul vero significato di questi esperimenti è quello di Onishi & Baillargeon (2005) che attesta che a partire dai 15 mesi i bambini sarebbero in grado di rappresentarsi le credenze dell'altro. Un ulteriore studio (Surian, Caldi, Sperber, 2007) abbasserebbe a 13 mesi l'età in cui i bambini sarebbero già in possesso di abilità di teoria della mente. E se non fosse davvero così? E se i bambini più che ad una teoria della mente si affidassero alla percezione di semplici regole comportamentali?

1.4 Discussione delle posizioni sull'*early mindreading*.

Nel più celebre degli esperimenti di Onishi & Baillargeon (2005) viene mostrato ad alcuni bambini un video che riproduce un attore intento a nascondere un oggetto in una scatola (quella verde). Dopo aver nascosto l'oggetto, viene posizionata una barriera di vetro che occlude la vista dell'attore, ma non quella del bambino. Subito dopo, l'oggetto viene spostato in un'altra scatola, quella gialla. L'esperimento si conclude con l'attore che si accinge a recuperare l'oggetto da una delle due scatole, sempre sotto gli occhi del bambino. La parte finale dell'esperimento, è quella che secondo Onishi & Baillargeon rivelerebbe le precoci abilità inferenziali del bambino. Quest'ultimo mostrerebbe, infatti, una

⁶ Questa batteria di test è un po' differente dalle altre. Viene valutata la capacità del bambino di distinguere la sua preferenza da quella di qualcun altro. In particolare il bambino viene messo di fronte a due pietanze. La prima è quella che lui preferisce. La seconda è quella preferita dall'esaminatore. Quest'ultimo fa comprendere la sua preferenza attraverso espressioni facciali ed espressioni di apprezzamento. I bambini di 18 mesi erano in grado di cogliere la preferenza (*desire*) dello sperimentatore.

tendenza a soffermare maggiormente lo sguardo quando l'attore cerca il giocattolo nella posizione reale. Sarebbe proprio questo *soffermarsi a guardare* ad indicare in maniera indiretta la consapevolezza dell'interiorità altrui da parte del bambino.

Queste le conclusioni a cui arrivano Onishi & Baillargeon:

These results suggest that 15-month-old infants already possess (at least in a rudimentary and implicit form) a representational theory of mind: They realize that others act on the basis of their beliefs and that these beliefs are representations that may or may not mirror reality. (Onishi & Baillargeon, 2005: 257)

Le affermazioni delle due studiose sono molto forti. Ciò che stanno dicendo è che già a 15 mesi i bambini sarebbero capaci di percepire gli altri come agenti intenzionali capaci di agire sulla base delle proprie credenze. Tutto questo grazie a dimostrazioni indirette da parte dei bambini esaminati.

È vero che affidare a dimostrazioni di consapevolezza più *dirette ed esplicite* la valutazione delle capacità inferenziali sarebbe troppo pericoloso. Si rischierebbe di non considerare come rappresentazionali, capacità che invece potrebbero esserlo e che potrebbero non apparire manifeste a causa della mancanza di altre abilità, assenti nei bambini di quell'età. Ciò che però è lecito chiedersi, è se ad una manifestazione così vaga e generica come la fissazione dello sguardo, davvero corrisponda la testimonianza di abilità di sicuro ben più complesse come quelle di meta-rappresentazione. È importante sottolineare che Onishi & Baillargeon non si limitano a constatare che in età così precoce sono presenti abilità di teoria della mente, ma arrivano a sostenere una tesi molto più forte e cioè che sia possibile spiegare tali abilità attraverso un meccanismo

metarappresentazionale dello stesso tipo di quello presente negli adulti. In disaccordo con le conclusioni a cui arrivano Onishi & Baillargeon, Perner & Ruffman (2005) ritengono che siano dei meccanismi di tipo percettivo-associativo ad essere i responsabili della fissazione dello sguardo dei bambini. Sarebbe a causa del fatto che i bambini si trovano di fronte al problema di dover formare una nuova associazione che fisserebbero per più tempo l'attore nell'atto di prendere la scatola vuota:

[...] we propose two explanations for the apparent early competence of infants that imply an evolutionary, innate bias for understanding the mind. Infants encode events and behavior the way they do because this encoding captures something useful about how people tend to act only because people are endowed with minds. Yet there is no need to assume an understanding on the infant's part that a mind mediates a particular behavior. Our first account of Onishi and Baillargeon's data is based on neuronal activation as babies process the events of the nonverbal false-belief task. Our suggestion is that babies create three-way actor-object-location associations. During the familiarization phase of the test, an adult actor watched by the infants last observes the object (water melon slice) in the yellow box under two conditions and in the green box under another two conditions. Neurons remember this information both in an active manner (through sustained firing in the prefrontal cortex) and in a latent manner (through altered firing thresholds in nonfrontal regions). If an association of elements "actor-object-yellow box" is still sustained in the frontal cortex when babies are exposed to the test stimuli, a consistent test combination will need less processing and, consequently, a shorter "looking" time than a new combination of elements (e.g., actor-object-green box). In the latter case, babies might need longer looking times because, when they

examine the new combination, they must form a new association. (Perner & Ruffman,2005:215)

Nessuna mentalizzazione, nessuna teoria della mente. Niente dell'esperimento di Onishi & Baillargeon farebbe pensare, secondo Perner & Ruffman alla necessità di introdurre il concetto di credenza per spiegare la fissazione dello sguardo.

La posizione di Onishi & Baillargeon, appartiene ad una corposa tradizione molto influente in scienze cognitive e in psicologia dello sviluppo, quella *innatista-mentalista*. A questa tradizione appartengono coloro che credono che le abilità meta-rappresentazionali siano presenti nel bambino molto precocemente, in virtù di una dotazione cognitiva presente fin dalla nascita:

Similar to other researchers, we assume that children are born with an abstract computational system that guides their interpretation of other's behavior. In this view, even young children appeal to others's mental states-goals, perceptions and beliefs-to make sense of their action; development involves primarily learning which states underlie which actions and not coming to understand that such states exist at all. (Onishi & Baillargeon, 2005: 257).

L'esperimento di Onishi & Baillargeon rende lecita la domanda: come è possibile giustificare il fallimento nei test di falsa credenza nei bambini prima dei quattro anni, se già a 15 mesi i bambini sarebbero capaci di pensiero inferenziale?

La risposta che fornisce Leslie et. al (1992;2005) risiede nel meccanismo del controllo inibitorio e dell'attenzione selettiva (SA). Nonostante il sistema credenze-desideri sia presente nei bambini già prima dei quattro anni, la non riuscita nei compiti di falsa credenza risiederebbe nella mancanza del sistema

inibitorio. I bambini, infatti, a quell'età sarebbero in possesso di una forma semplificata di ToMM. Quest'ultima non permetterebbe loro di immaginare con facilità situazioni differenti da quella attuale. Di conseguenza, per essere capaci di comprendere un'azione che non rispecchia la realtà delle cose (e dunque superare i test di falsa credenza), sarebbe necessario far intervenire un sistema in grado di inibire il funzionamento di questo meccanismo semplificato e permettere un ragionamento controfattuale:

Instead, we argue that the critical factor is the pattern of inhibitory processing in a belief-desire task that requires shifts in the target of attention. [...] The failure of children on false-belief tasks is the result of a failure in the selection process to effectively inhibit the true-belief attribution. To the extent that this is true, there are a number of important theoretical consequences. First, what has been the single most important objection to the ToMM account is removed. Second, we will have advanced our understanding of a domain-specialized learning process. Attributing beliefs, with false or with true contents, is unique to the theory of mind domain, yet children may come to learn about false beliefs by way of increased powers of selection by inhibition. Third, we will extend our understanding of the role of inhibition in reasoning and development. Thus, although belief attribution is domain specific, selection by inhibition may operate in a variety of domains. In this case, belief-desire reasoning and its development provide an avenue for studying inter-relations between domain specific and domain general processes of learning and reasoning. (Leslie, German, Polizzi, 2005: 41)

Quello di Leslie è uno dei possibili modi in cui un innatista-mentalista risolverebbe la questione. Il fatto che prima dei quattro anni i bambini non siano capaci di risolvere compiti di falsa credenza, non significa che nel bambino non siano presenti abilità di *mindreading* già a 15 mesi.

Esattamente come per la proposta di Onishi & Baillargeon, Ruffman & Perner (2005) in un articolo di risposta a Leslie, sottolineano l'inappropriatezza della sua proposta. La questione, esattamente come nel caso delle abilità inferenziali delle antropomorfe, può essere risolta in maniera meno dispendiosa.

Così scrivono:

Leslie questions why infants would evolve behavior rules. Our points are that such rules are potentially used by primates with which humans share a genetic ancestry, that such ancestral abilities also form core theories in other areas that are then modified by means of language, or that infants' sophisticated statistical learning abilities would also provide ample means for acquiring such rules. Indeed, the parallels to language acquisition are striking; beliefs that syntax must be innate have been tempered by evidence that infants' statistical learning abilities (which might themselves be innate) permit learning about at least some aspects of syntax. Leslie also argues that 3-year-olds do not use behavior rules in the traditional false-belief task but default to answering in terms of reality. This leaves the question of why their looking to the empty location in implicit tasks does not also default to reality. Instead, such looking is consistent with the use of a rule. It is perfectly plausible that children use this rule in an implicit task but when asked explicitly they use a different strategy. (Ruffman & Perner, 2005: 462)

Secondo Ruffman & Perner (2005), sarebbe inoltre azzardato parlare di “comprensione di credenze”. La sperimentazione attuale, secondo i due studiosi, non ci direbbe nulla in favore di questa possibilità:

Leslie suggests that younger children (below 2 yrs 11 mth) fail implicit measures because of the verbal demands of the implicit tasks. Yet these tasks follow the same sequence of events as in Onishi and Baillargeon’s study. Even without understanding the narrative, the visual details of the events should enable correct eye gaze, provided that infants understand the eye gaze prompt (e.g. ‘I wonder where he’ll look?’). Importantly, 2-year-olds do look correctly in response to this identical prompt in other social understanding tasks [15], and ironically also in a study cited by Leslie (Waskett et al., unpublished), demonstrating the verbal demands are within their grasp. In our view, current data indicate that infants understand much about behavior but whether it includes an understanding of belief is still a wide open question. (Ruffman & Perner, 2005: 463)

Una posizione molto interessante, all’interno di questo dibattito sulle abilità inferenziali nei bambini, è quella di Carruthers (2013). Vicino alle posizioni innatiste, quest’ultimo ritiene che la spiegazione percettivo-associativa, non sia sufficiente a giustificare le capacità inferenziali dei bambini. La sua posizione è molto vicina, per sua stessa ammissione, a quella di Leslie, che abbiamo descritto in precedenza. Carruthers ritiene infatti che esista un meccanismo dominio-specifico presente già nella prima fase dell’infanzia che provvede a fornire al bambino gli elementi necessari alla comprensione degli stati mentali altrui. Questo meccanismo che diventa sempre più efficiente durante la crescita del bambino, fornisce già da subito l’abilità di ragionare in termini di falsa credenza. Ciò che permette però al bambino di sviluppare pienamente il pensiero

inferenziale, e ciò che dunque segna la differenza tra l'*early mindreading* e il tardo *mindreading*, è l'interazione tra questo sistema principale (strettamente inferenziale) ed altri meccanismi esecutivo-attenzionali:

The domain-specific mechanism implicated in mindreading competence is functional early in infancy. (I shall tentatively suggest: by the middle of the first year of life.) This provides infants with the concepts and core knowledge necessary to represent the mental states of other agents, of all basic types (including beliefs that are false and appearances that are misleading). While the operations of this system probably become more streamlined and efficient with age, its representational capacities do not alter in any fundamental way. Rather, what matures over time are the quality and extent of the interactions between this system and executive, attentional, and planning mechanisms. (Carruthers, 2013: 142)

Se da un lato questa ipotesi sembra orientarsi verso posizioni di stampo innatista, dall'altro sembra lasciare ampio spazio alle regole comportamentali e contestuali. Vediamo, però, in che modo. Secondo Carruthers, solo nel momento in cui si ammette l'esistenza di un meccanismo innato che contiene in sé la capacità di svolgere compiti di falsa credenza, si può dire che le euristiche comportamentali abbiano un ruolo nella formazione del pensiero inferenziale. Se non si ammette l'esistenza di questo "innately channeled body core knowledge" è impossibile pensare che i meccanismi associativo-percettivi abbiano un ruolo nella comprensione dell'intenzionalità altrui:

It is important to note, however, that the proposal of an early-developing core mindreading system is not intended to rule out a subsequent role for learning. On the contrary, it is likely that the system is designed to enrich itself as development proceeds, acquiring new ways of inferring people's

mental states from behavioral or contextual cues, for example. There is also likely to be expansion in the range of mental states that the system is capable of representing, adding new non-basic concepts to its stock of representations (such as concepts of socially constructed types of emotion). What is present at the outset is just a capacity to attribute propositional representations of various basic types, including belief. (Carruthers,2013: 142)

Ciò che in definitiva Carruthers critica degli approcci fondati sulla lettura del comportamento è che essi non spiegano in che modo queste regole comportamentali possano essere apprese nel primo anno di vita, rendendo questo tipo di approcci implausibili. Ciò che invece può essere considerato plausibile, anche da un punto di vista evuzionistico, è che grazie a pressioni selettive favorevoli si sia evoluto un sistema di “core knowledge” capace di fungere da base per l’acquisizione di più complesse competenze inferenziali:

The infant-mindreading hypothesis, in contrast, postulates an innately channeled body of core knowledge, or an innately structured processing mechanism (or both), with an internal structure that approximates a simple theory of mind. The explanatory burden, then, is an evolutionary one: it needs to be shown that there were sufficient adaptive pressures among our ancestors for such a mechanism to evolve. There is now an extensive body of work suggesting that this is indeed the case. The gains provided by such a mechanism might derive from enabling so-called ‘Machiavellian intelligence’ (Byrne and Whiten, 1988, 1997), from facilitating larger group sizes (Dunbar, 1998), from enabling distinctively human forms of cooperation and collaboration (Richerson and Boyd, 2005; Hrdy, 2009), or from any combination of these. (Carruthers,2013: 151)

Formulare un modello teorico chiamando in causa la teoria dell'evoluzione naturale rende ancora più chiara, per Carruthers, la necessità di ribadire che non esiste alcuna differenza qualitativa tra l'early mindreading e le più tarde capacità metarappresentazionali. Questa presa di posizione, oltre a sancire come abbiamo visto, un distacco definitivo dagli approcci *behaviour-rules based*, differisce da un altro tipo di approccio che postula l'esistenza di due differenti sistemi di apprendimento inferenziale: uno rigido, automatico e veloce e l'altro più lento e controllato. Stiamo parlando dei *dual systems* alla base dell'ipotesi di Apperly & Butterfill (2013).

In disaccordo con Carruthers, questi ultimi ritengono che esista una differenza qualitativa tra i due sistemi inferenziali. Nella prima infanzia, è presente solo il primo sistema, quello inflessibile. Quest'ultimo essendo automatico non permette la rappresentazione di credenze, ma garantisce velocità ed efficienza. Solo in seguito, attraverso lo sviluppo del linguaggio e la maturazione delle funzioni esecutive, è possibile l'utilizzo di un sistema più flessibile e di conseguenza più complesso. È importante notare che i due sistemi coesistono nell'uomo adulto e si attivano in base alle esigenze a cui devono far fronte:

Human infants and some nonhuman animals are able to solve some theory-of mind tasks by virtue of having one or more systems that are cognitively efficient but limited and inflexible. These limitations are only overcome years later when human children acquire psychological concepts such as belief and desire; this development occurs gradually (e.g., Apperly & Robinson, 2003), appears to be related to the development of language (Astington & Baird, 2005) and executive function (Perner & Lang, 1999; Sabbagh, 2006), and may be facilitated by explicit training (Clements et al., 2000; Slaughter & Gopnik, 1996). This equips children with a new system for theory-of-mind reasoning that is highly flexible but cognitively inefficient. In human adults, both systems exist in parallel. The cognitively efficient system

plays a central role in guiding online social interaction and communication. The cognitively flexible system enables adults to engage in top-down guidance of social interaction (such as anticipating what the audience of a lecture might know or working out how one misjudged the audience afterward) and in explicit reasoning about the causes and justifications of mental states (as in everyday practical reasoning or jurisprudence). (Apperly & Butterfill, 2009: 966)

Questa ipotesi, ad un primo sguardo, potrebbe sembrare vicina agli approcci *behaviour-rules based* che abbiamo considerato in precedenza. In fondo si potrebbe dire che quello che Apperly & Butterfill definiscono il più primitivo sistema inferenziale, quello automatico e rigido, non sia così tanto differente dalle ipotesi basate sull'associazionismo percettivo. A ben guardare, però, Apperly & Butterfill sembrano volersi allontanare da questa prospettiva quando introducono la definizione di "minimal theory of mind":

[...] theory of mind abilities are widespread, occurring not only in human adults but also in infants, chimpanzees and scrub-jays at least. By contrast full-blown theory of mind cognition may be comparatively rare thanks to the conceptual and cognitive demands associated with representing beliefs and other propositional attitudes as such. If so, it is useful to identify what else someone could represent that would enable her to track, in a limited but useful range of situations, perceptions, knowledge states and beliefs, including false beliefs. To answer this question we construed a minimal theory of mind. [...] Importantly, we do not assume that minimal theory of mind develops into full-blown theory of mind in humans. It may instead

remain distinct, supporting cognitively efficient theory of mind across the lifespan. (Apperly & Butterfill, 2013: 629)

Quello che stanno dicendo è che è plausibile che esista uno stadio intermedio tra la semplice lettura del comportamento e le abilità inferenziali più complesse che permettono la rappresentazione di credenze. La “minimal theory of mind” è più sofisticata delle behaviour rules, ma allo stesso tempo non può essere considerata una vera e propria teoria della mente. Questa assunzione può essere fatta sulla base di una distinzione necessaria secondo Apperly e Butterfill: quella tra *theory of mind ability* e *theory of mind cognition*. La prima permette il riconoscimento delle percezioni e delle credenze altrui, la seconda permette la rappresentazione di stati mentali. Per dare supporto a questa distinzione hanno arricchito la “minimal theory of mind” con due nozioni alternative a quella di percezione e di credenza: quella di *encountering* e quella di *registration*. Due nozioni meno forti e con una connotazione mentalistica più debole. Entrambe giocano il ruolo di intermediari nella costruzione di percezione e credenze:

Encountering and registration serve as non-representational proxies for perception and belief in this sense: in a limited but useful range of everyday circumstances, agents perceive an object just when they encounter it and they believe that an object has a given property just when they register it having that property. But encountering differs from perception in not constitutively involving appearances, modalities or the possibility of illusion and in being constitutively linked to reasons, knowledge or informational states. Similarly, registration differs from belief in lacking representational content and the possibility of embedding, in not interacting in arbitrarily complex ways with other states, and in the simplicity of its parameter-

setting role in causing actions. These differences ensure that representing encountering or registration need to not involve meeting conceptual and cognitive demands associated with representing perceptions, beliefs and other propositional attitudes as such. (Apperly & Butterfill,2013: 629)

Ci si potrebbe chiedere a questo punto se la *minimal theory of mind* si traduca concretamente in una tappa intermedia dello sviluppo delle abilità inferenziali o se si riduca semplicemente all'utilizzo di una nuova terminologia per identificare una categoria la cui esistenza non è supportata da dati empirici. L'idea che debbano esistere due distinti sistemi, di cui uno garantisce efficienza e l'altro flessibilità, non convince Christensen & Michael (2013) che propongono una visione del *mindreading* fondata sulla cooperazione di più sistemi:

The most problematic aspect of Butterfill and Apperly's restricted empirical scope is that the overall pattern of evidence raises issues that their account does not appear well-equipped to deal with. Taken as a whole, the evidence reveals striking flexibility in infant belief representation abilities, whereas the mindreading principles described above are notably inflexible. It is not obvious how they can be straightforwardly extended to encompass the representation of beliefs that involved diverse properties, beliefs acquired through communication, and so on. We'll argue that explaining this flexibility requires an approach that is in key respects fundamentally different. A cooperative multi-system architecture is better able to explain infant belief representation than a parallel architecture, and causal representation, schemas and models provide a more promising basis for flexible belief representation than does a rule-based approach of the kind described by Butterfill and Apperly. (Christensen & Michael,2013: 21)

Christiansen & Michael trovano implausibile l'approccio di Apperly & Butterfill per diverse ragioni. In primo luogo, non credono sia possibile considerare la flessibilità come un'abilità che si sviluppa solo quando viene attivato il secondo sistema, quello più complesso. Ci sarebbero infatti evidenze chiare in favore del comportamento flessibile nei bambini molto piccoli (in cui ancora il secondo sistema non è attivo). Secondo Christiansen & Michael, infatti, non possono esistere comportamenti inferenziali se non si è capaci di agire in maniera flessibile. In secondo luogo, i due studiosi ritengono che non sia corretto ragionare nei termini del *rules-based approach* (principio che regola il sistema automatico-inflessibile di Apperly & Butterfill) che utilizza il criterio del metodo statistico. Propongono, invece, un'architettura fondata sulla cooperazione tra più sistemi che interagiscono tra loro e si scambiano informazioni utilizzando il criterio della *causal representation*. Secondo Christiansen & Michael, la *minimal theory of mind* essendo automatica e rigida, non può permettere una rappresentazione il cui contenuto posseda la proprietà di causa-effetto. Dunque secondo Christiansen & Michael il modello di Apperly & Butterfill non può funzionare perché non è possibile concepire un sistema inferenziale in cui non ci sia legame causale tra percezione e azione. *Encountering* e *registration*, le due componenti della *minimal theory of mind*, non sarebbero dotate di proprietà causali:

[...] they are defined in a way that includes no causal connection between perception and action. Encountering and registration are simply specified as necessary conditions for successful goal-directed action. This is very different and much weaker than representing perception as providing information for action. The latter is causal, the former is not. (Christiansen & Michael, 2013: 25)

Quello che propongono è dunque un'architettura fondata sulla cooperazione tra sistemi, in cui le regolarità comportamentali sono assemblate in modelli e schemi tenendo conto delle loro proprietà causali. Solo così è possibile non rinunciare all'efficienza in favore della flessibilità, necessaria nelle abilità di mindreading.

Ci si potrebbe chiedere a questo punto se le critiche mosse ad Apperly & Butterfill riguardo alla presunta non causalità attribuita alla loro *minimal theory of mind* siano lecite. E più in generale se sia davvero possibile che un sistema seppur automatico possa elaborare informazioni in modo da non rispettare la nozione di causalità⁷.

Fin adesso abbiamo ripercorso i modi in cui viene affrontata la questione dell'attribuzione di credenze.

Abbiamo visto che mentre alcuni ritengono che gli esperimenti forniti da Onishi & Baillargeon siano la testimonianza che già a 15 mesi i bambini possiedono una teoria della mente, altri ritengono che la sola lettura del comportamento sia sufficiente a spiegare i risultati degli esperimenti effettuati.

La posizione di Cecilia Heyes (2014) merita un discorso a parte. Critica nei confronti delle posizioni di stampo innatista, ritiene errate e ingiustificabili le conclusioni a cui arrivano i sostenitori di questo approccio. Ritiene infatti che in assenza di linguaggio (ed è il caso analizzato da Onishi & Baillargeon) non sia così facile stabilire se il bambino attribuisce credenze sulla base di una teoria della mente o sulla base di abilità di lettura del comportamento:

It is notoriously difficult to specify exactly what it would be 'really' to attribute mental states, and how this differs from the use of behaviour rules. However, a common intuition is that real mentalizing involves going beyond

⁷Il dibattito tra Apperly & Butterfill e Christensen & Michael sarà ripreso in maniera più approfondita nel paragrafo 3.2

observables relations between circumstances and behavior, and representing these regularities as indicative of unobservable (theoretical or subjective) causal states with intentional content; states such as beliefs and desires. How might mentalizing in this sense be distinguished empirically from the use of behavioural rules alone, in tests that do not require language? (Heyes, 2014: 10)

La Heyes ritiene infatti che i dati empirici finora disponibili, non rivelano nulla in favore della versione innatista-modularista. Le sue critiche si rivolgono in particolare all'idea che esista un meccanismo dominio-specifico legato all'interpretazione di credenze:

[...]my analysis suggests that dedicated mentalizing processes may not be necessary; that the same jobs can be done just as effectively by domain-general processes such as those involved in automatic attentional orienting and spatial coding of stimuli and responses. There is no reason to suppose that these domain-general processes evolved 'for' the tracking of mental states, that their development is especially dependent on interaction with other agents, or that they operate selectively or in a distinctive way in social contexts. However, I have called them 'sub-mentalizing' rather than, for example, 'pseudo- mentalizing' because I think these, and other domain-general processes, provide substitutes and substrates for mentalizing in everyday life. (Heyes,2013: 20-21)

La sua idea è dunque che non esiste alcun processo di mentalizzazione responsabile del nostro modo di ragionare in maniera inferenziale. Quella che lei definisce *sub-mentalization* non è altro che la maniera di identificare un processo generale per dominio che opera nelle inferenze mentalistiche esattamente come in qualunque altra nostra attività. Con queste premesse è ovvia la sua posizione nei confronti degli esperimenti fondati sulla violazione dell'aspettativa (*violation of expectation*) e sullo sguardo anticipatorio (*anticipatory looking*). Ritiene infatti che questi esperimenti lungi dal dimostrare l'esistenza di una teoria della mente implicita, muovono in favore della *low-level novelty* da lei formulata.

I argue that the results of infant FB studies are due to low-level novelty. In many cases, the novelty depends only on the observable properties of the test stimuli; infants look longer at test events that, when compared with events encoded earlier in the experiment, display new spatiotemporal relations among colours, shapes and movements (perceptual novelty). In other cases, the novelty depends on imagination as well as the manifest characteristics of the test stimuli; infants look longer at test events that, when compared with events encoded earlier in the experiment, are remembered or imagined to have new spatiotemporal relations among colours, shapes and movements (imaginal novelty). The low-level novelty account assumes that domain-general processes of perception, attention, motivation, learning and memory are solely responsible for encoding the events that infants witness in FB experiments and for modulating the infant's looking behaviour. It is not rooted in particular theories of how these domain-general processes operate. (Heyes,2014: 2)

Capitolo 2

Flessibilità cognitiva, apprendimento associativo e mindreading

L'analisi della letteratura sull'*early mindreading*, ci ha portato a concludere che né l'approccio innatista né quello fondato su regole comportamentali sono sufficienti a rendere conto del modo in cui i bambini riescono a prevedere e interpretare in maniera precoce il comportamento altrui. I due principali approcci alla questione dell'attribuzione precoce di credenze si rivelano da un lato insufficienti nella spiegazione del fenomeno su cui intendono far luce, dall'altro mostrano, a mio avviso, di avere in comune un'assunzione che rimane sostanzialmente implicita: entrambi presuppongono una forma di flessibilità. Quest'ultima, altrimenti chiamata "set shifting, mental flexibility, or mental set shifting and closely linked to creativity" (Diamond,2013), è una caratteristica essenziale nell'elaborazione di qualsiasi tipo di informazione che potrebbe giocare un ruolo chiave anche nella costruzione del *mindreading*.

Comprendere in che modo l'agire flessibile è possibile significa riflettere sul tipo di architettura cognitiva che può ragionevolmente essere alla base di tali comportamenti. La letteratura in merito si divide sostanzialmente in coloro che credono che sia un'architettura cognitiva perlopiù innata a permettere la flessibilità dei comportamenti ed in coloro che credono invece che la flessibilità sia possibile con un'architettura cognitiva perlopiù acquisita. Nello specifico la letteratura è divisa tra coloro che credono che alla base dei comportamenti flessibili ci siano meccanismi cognitivi di tipo automatico, e più in particolare credono che la risposta risieda in una organizzazione di tipo massivamente modulare, e coloro che invece credono che sia una struttura di tipo gerarchico-associativo a garantire grazie a meccanismi controllati la flessibilità dei comportamenti.

2.1 Flessibilità cognitiva e modularità massiva

Tra coloro che credono che la flessibilità cognitiva sia garantita da un'architettura cognitiva innata complessa e articolata, troviamo i sostenitori della modularità massiva.

L'idea che quest'ultima possa garantire comportamenti flessibili è intuitivamente poco plausibile. Basta pensare all'impermeabilità tra moduli, una delle peculiarità della modularità massiva, per trovare poco probabile che una architettura cognitiva di questo tipo possa garantire qualsiasi forma di flessibilità. Peter Carruthers in *The Architecture of the mind* (2006) si occupa proprio di risolvere questa, secondo lui solo apparente, incompatibilità provando a sostenere la tesi secondo cui sarebbe invece proprio una architettura di tipo massivamente modulare a garantire e rendere possibile la flessibilità dei comportamenti.

Dopo aver fornito differenti definizioni di flessibilità cognitiva ne individua differenti tipi (*flexibility of content, stimulus independence, flexibility of reasoning, e flexibility of decision making*) fino ad arrivare a definirla come la più generale capacità, estremamente sviluppata nell'uomo, di formulare pensieri che non hanno alcuna relazione con le circostanze attuali. ("We routinely entertain thoughts, and whole sequences of thought, that bear no relation to our current physical or social circumstances").

La sfida più grande che deve superare la modularità massiva è sicuramente la flessibilità del contenuto (*content flexibility*). Questo tipo di flessibilità è del tutto inconciliabile con l'impermeabilità tra moduli. L'esempio di Machery (2008) rende chiara questa incompatibilità. In che modo riusciamo ad attribuire un significato alla frase *My brother is a cheater?*

Brother e *cheater* appartengono a due moduli differenti, il *cheater detection module* e il *kinship detection module*⁸. In che modo questi concetti possono essere processati in modo da formare un unico concetto (cheating brother)? Come facciamo a formare concetti complessi? In che modo un modulo può processare la rappresentazione del suo contenuto? Come si combinano pensieri provenienti da domini diversi? In che modo la flessibilità può dunque essere garantita da un'architettura cognitiva di tipo massivamente modulare?

Carruthers è perfettamente cosciente dei problemi che pone la modularità massiva e per provare a risolvere le criticità insite nel suo modello di architettura cognitiva, propone l'introduzione di un ulteriore sistema di processazione il

Practical reasoning system:

I can be thinking about thought one moment, horse the next, and then a landslide the next; and I can then wonder what led me to think about thoughts, horses, and falling stones- thereby combining into a single thought concepts drawn from the domains of folk-psychology, folk biology, and folk physics. And likewise, for any set of conceptual modules that you care to mention. How is this possible, unless there is some a-modular central arena in which the contents of conceptual modules can be received and recombined, further inferences drawn from the results, and so forth? This is yet another challenge to massive modularity. It might be replied that I have, over the previous two chapters, committed myself to the existence of just such a central arena, namely the practical reasoning system. I have suggested that this system is capable of receiving any desire and any belief as input. So why shouldn't it be this system that has the power to combine

⁸ Questa è una tesi ampiamente condivisa e Machery (2008) decide di adottarla nel caso specifico dei due moduli da lui presi ad esempio. Si può assumere però che l'argomento valga in generale: in caso si abbiano idee differenti su quali moduli esistano, è sufficiente sostituire questi con altri due domini qualsiasi che si assume siano modulari.

and recombine concepts drawn from disparate domains? (Carruthers,2006: 226)

Ipotizzando l'esistenza di un sistema che processa concetti che provengono da domini separati viene forse risolta la questione dell'incomunicabilità tra moduli. Resta il fatto, però, che pur accettando l'esistenza del *practical reasoning system*, rimane il problema di spiegare in che modo questi contenuti siano integrati all'interno del sistema.

Dovremmo forse ipotizzare l'esistenza di un altro sistema che a sua volta integra i contenuti all'interno del *practical reasoning system*? Si ricadrebbe in una regressione all'infinito. Carruthers, probabilmente consapevole delle criticità insite nel suo modello, prova a risolvere la questione ipotizzando altri due possibili candidati atti ricoprire il ruolo di "integratori di contenuti" insieme al *practical reasoning system*: Il linguaggio, e il mindreading. Vediamo in che modo:

I have suggested that if the mind were massively modular, then we should expect there to be some limits on the ways in which concepts can be combined and integrated with one another [...]. Yet we have reason to think that there are no such limits on the flexibility of the human mind. So we have a problem: how can such flexibility of content arise in a massively modular mind? I have been suggesting that there are reasons of a general sort for thinking that it is the language production module that performs such a role, initially in the service of speech efficiency. (Carruthers,2006: 242-243)

Il candidato più plausibile sembra essere dunque proprio il linguaggio. Per comprendere perché quest'ipotesi sembra la più plausibile, dobbiamo capire le

ragioni che conducono Carruthers a escludere il mindreading (l'altro candidato) dall'occupare il ruolo di "integratore di contenuti". Vediamo come viene giustificata l'esclusione:

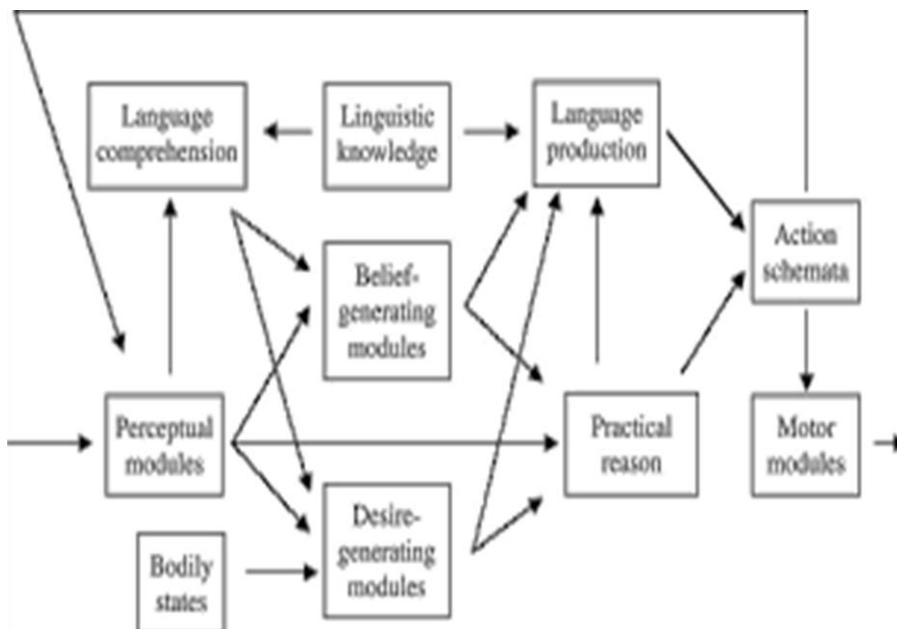
Attributing mental states to other people on the basis of behavioral cues, and/or predicting people's behavior from mental states previously attributed to them, is what the mind-reading system is primarily about. Combining concepts and propositions, and drawing arbitrary inferences from them, would seem to be a distinct set of functions entirely. In which case we should expect that the latter functions will be carried out in one or more distinct modules. For there is no reason to think that they will come 'for free' with mind-reading functions. In contrast, there is good reason to think that the content-combining functions will come for free with the language module, given the constraint of speed of sentence-production. (Carruthers, 2006: 248-249)

E ancora continua:

In the case of language, in contrast, the demands of swift and efficient communication would have created a significant selection pressure for intermodular integration, allowing the outputs of distinct central modules concerning the same object or event to be combined into a single spoken sentence. So instead of saying separately, 'The object is near a short wall', and, 'The object is near a red wall', one can say much more succinctly, 'The object is near a short red wall' (given that the short wall in question is the red wall, of course). (Carruthers, 2006: 249)

In sintesi Carruthers confrontando strutture e funzioni dei due moduli, quello del linguaggio e quello del mindreading, si rende conto che il linguaggio si presta in maniera migliore a ricoprire il ruolo di “integratore di informazioni”, in questo modo esclude il mindreading.

Nella figura in basso è rappresentata in sintesi “l’architettura della mente” così come ipotizzata da Carruthers, in cui il modulo di “language production” e il “practical reasoning system” lavorano per integrare le informazioni provenienti dai differenti domini. Ipotizzando la cooperazione tra questi sistemi è davvero possibile un’integrazione dei contenuti e quindi una sua conseguente flessibilità?



(Carruthers,2006: 233)

Un modo forse più convincente di concepire la flessibilità è fornito da Dan Sperber che riformula l'idea stessa di modulo cognitivo. Vediamo in che modo:

Most innate cognitive modules are domain-specific learning mechanisms (“learning instincts” [Marler 1991], or “module templates” [Sperber 1994]) that generate the working modules of acquired cognitive competence.[...] With many innate modules being learning modules generating further modules, with brain areas ready to modularize, one may envisage that the human mind is characterised not only by massive modularity, but also by teeming modularity. A great many highly specialized procedures, the size, say, of a specific concept or even of a particular inference rule, may be modular in the intended sense. That is, there may be a plethora of distinct biological devices emerging on some innate basis in the course of cognitive development, and functioning with a certain degree of autonomy in cognitive activity (a similar view, based on an analogy between cognitive modules and enzymes, is developed by Clark Barrett, forthcoming). I hope these remarks help understand how a massively modular mind may indeed be flexible, even if the detailed ways in which such flexibility is achieved obviously are a matter for empirical research. (Sperber, 2005)

È affidando ai meccanismi biologici (innati) un certo grado di modificabilità che Sperber risolve in qualche modo la questione del processamento delle informazioni provenienti da moduli differenti. In sostanza l'idea di modularità massiva sarebbe compatibile con l'idea stessa di flessibilità solamente riducendo il peso della biologia in favore di una visione che assegna al modulo una seppur parziale capacità di plasmarsi grazie agli stimoli provenienti dall'ambiente⁹.

⁹ A proposito di rivisitazione dell'idea di modulo si veda Karmiloff-Smith(1992).Anche in questo caso assistiamo, in maniera più convincente, ad una ridefinizione dell'idea di modulo fondata su un ridimensionamento del peso della biologia in favore dell'esperienza.

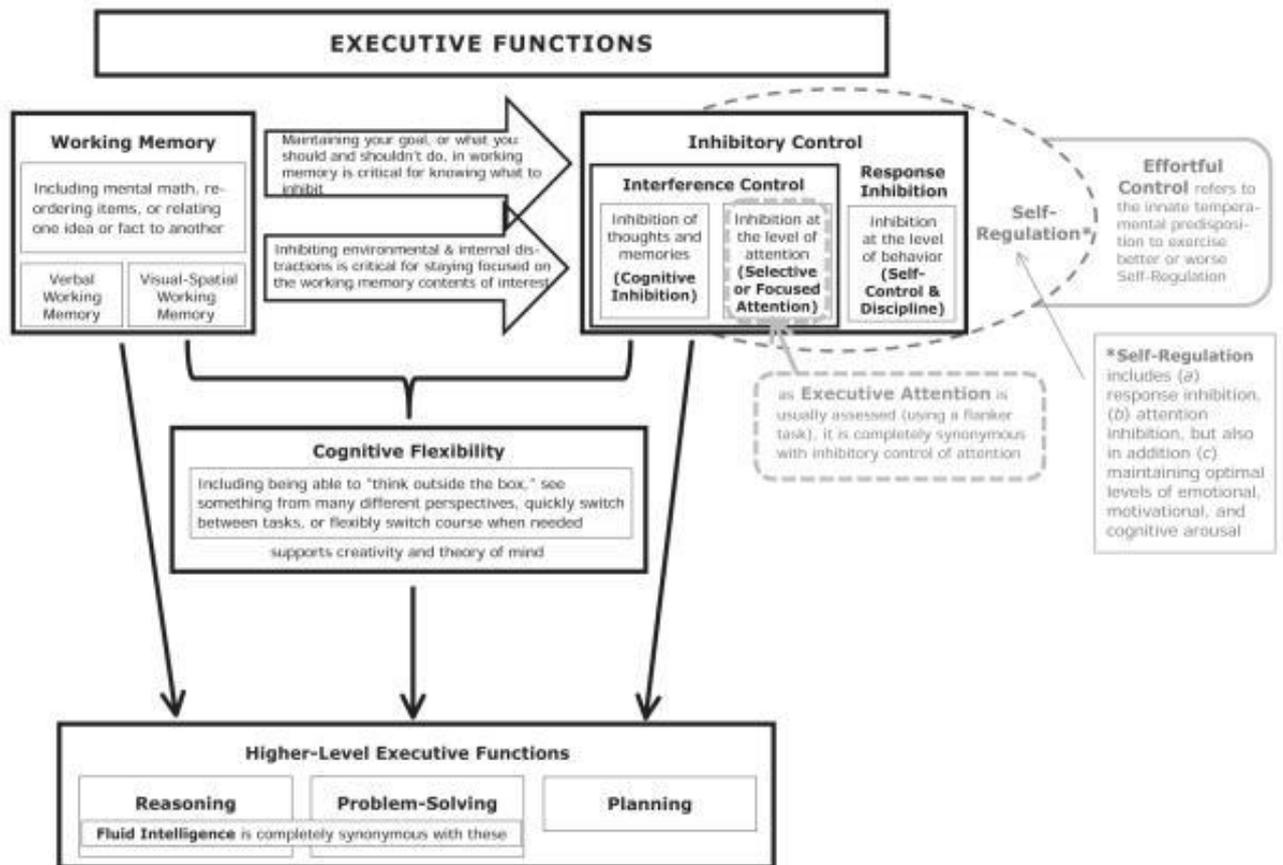
La visione modularista appare problematica, su molteplici aspetti. Prima di tutto ci si potrebbe chiedere se è davvero legittimo assecondare la tendenza a creare entità la cui esistenza risulta dubbia e valida su un piano esclusivamente speculativo. Ci si potrebbe chiedere insomma se ha davvero senso creare una categoria ad hoc come per esempio quella del *Practical reasoning system* per giustificare il bisogno di elaborare le informazioni che provengono da domini differenti.

2.2 Flessibilità cognitiva e funzioni esecutive

A prescindere da quali meccanismi rendano possibile la flessibilità cognitiva, ciò che è accettato largamente è che la flessibilità sia una parte di un sistema di funzioni: le funzioni esecutive.

There is general agreement that there are three core EFs (e.g., Lehto et al. 2003, Miyake et al. 2000): inhibition [inhibitory control, including self-control (behavioral inhibition) and interference control (selective attention and cognitive inhibition)], working memory (WM), and cognitive flexibility (Diamond,2013: 136)

In sostanza è necessario inserire la flessibilità cognitiva all'interno di un sistema di questo tipo in cui la flessibilità sia una parte della faccenda, inscindibile dalle altre (memoria di lavoro e controllo inibitorio):



(Diamond, 2013)

Come possiamo osservare dal grafico, la flessibilità cognitiva è quella abilità che permette di “pensare fuori dagli schemi”, analizzare differenti prospettive e scegliere quella attraverso la quale è possibile arrivare allo scopo prefissato adattandosi alle nuove situazioni. Inoltre, viste queste caratteristiche, è ovvio il ruolo che quest’ultima può ricoprire in compiti di teoria della mente. In che modo le due abilità siano legate appare poco chiaro nella letteratura ad oggi consultata. Ma questo lo vedremo più avanti.

2.3 Flessibilità cognitiva e meccanismi automatici

Un altro possibile modo di approcciarsi alla flessibilità è quello di considerarla dipendente da meccanismi di tipo automatico. Quest'idea può sembrare ad un primo sguardo contro-intuitiva. Come fa notare Hassin (2009), di solito la letteratura tende a far coincidere flessibile con controllato e rigido con automatico. Ciò che vuole fare lo studioso è proprio sciogliere queste coppie concettuali e sfatare dunque questa convinzione comune. In che modo l'agire flessibile potrebbe non essere cosciente? Per rispondere a questa domanda analizziamo una delle possibili definizioni di flessibilità:

One way to conceive of flexibility is as the “capacity for ready adaptation to various purposes or conditions” (Oxford English Dictionary, italics added). Flexibility of this kind may have cognitive components (e.g., realizing that a rule that governed the environment has changed), behavioral components (e.g., altering one’s behaviors accordingly), and affective ones (e.g., not resenting the new rule just because it is new. (Hassin, Bargh, Zimerman 2009: 3)

La flessibilità è definita dunque come la capacità di “adattarsi a nuove condizioni”. In che modo questa necessità può essere soddisfatta da un meccanismo automatico? La proposta di Hassin Bargh e Zimerman, si fonda principalmente su alcune evidenze empiriche e sulla convinzione che nell'adattarsi a nuove condizioni i meccanismi coscienti non abbiano l'importanza che si è soliti attribuirgli. In questa visione, il ruolo assunto dai meccanismi di tipo cosciente nell'agire flessibile viene ridimensionato perché ritenuto poco pertinente e poco plausibile:

A significant proportion of human behavior is determined by non-conscious goal pursuits. This assertion is easily derived from two well-established and highly consensual observations about human nature. First, much of human behavior is purposeful, or goal directed. Our goals range from very trivial (e.g., to make a cup of coffee) and a little less so (e.g., to get to work), through more complex (e.g., to write an interesting chapter) to extremely difficult ones (e.g., to be a good parent). It is not completely unlikely that goals direct behavior at virtually every moment of our lives. Second, our consciousness is very – but very – limited in its processing resources. Memorize simple cooking instructions, count the number of knives you put on the table, or just think a simple thought – and your conscious capacity drops substantially. (Hassin, Aarts, Eitam, Custers, Kleiman, 2009: 2)

Una caratteristica di questi modelli che gioca un ruolo fondamentale nel garantire la flessibilità è la dipendenza dal contesto:

Although at first sight ‘flexibility’ seems to be an innocuous concept, it is truly complex and -appropriately- context dependent (Hassin, Bargh, Zimerman, 2009: 3)

Ogni comportamento, infatti, è innescato ed è reso possibile da ed in quel preciso contesto:

[...] the context of meeting an attractive colleague may instigate the goal of intimacy, which may bring about a certain way of talking that has proved in the past to be an effective means for attaining this goal. This chain of events (i.e., a context activates a goal that activates certain means) may occur outside of awareness, without a conscious decision, and sometimes even

despite one's conscious intentions. (Hassin, Aarts, Eitam, Custers, Kleiman, 2009: 4)

Il meccanismo che si attiva è simile a quello che descrive Smith (2000) a proposito dell'apprendimento associativo. Quest'ultimo, proprio come il processo descritto da Hassin, è costruito storicamente e per funzionare e produrre output efficaci e pertinenti deve avvenire all'interno di contesti specifici in cui sia possibile costruire catene e legami causali:

[...] the uniqueness of the outcome depends exquisitely on the particulars in the history of the processes and the cascading consequences within that history. In other words, what matters is development. (Smith, 2000: 172)

È da qui che deriva infatti una delle caratteristiche dell'apprendimento associativo:

Associative learning operates in particular contexts, as parts of particular causal chains and particular natural histories, and yields specific outcomes (Smith, 2000: 173)

Modelli di questo tipo si fondano in larga parte sull'idea che un contesto già noto permette l'attivazione di pattern comportamentali riferiti all'esperienza di quel contesto che si attiveranno ogni qual volta ci troveremo di fronte a quella stessa situazione. Cosa succede se ci troviamo invece di fronte a situazioni "nuove" o

comunque non assimilabili alle precedenti? Che tipo di meccanismo agirebbe di fronte a situazioni di questo tipo? È possibile fare a meno dei processi coscienti anche quando non possono entrare in gioco le *routine*?

So how do we go beyond pre-existing routines in non-conscious goal pursuit? The traditional answer would be that we do not: Non-conscious goal pursuit, like every other automatic process, is limited to circumstances in which pre-existing routines could be successfully applied. If they cannot be successfully applied, then non-conscious goal pursuit is bound to fail and conscious processes would be called to the fore. While tempting, this suggestion is psychologically improbable. Given the scarcity of conscious mental resources on the one hand, and the dynamic nature of the world we inhabit on the other, it seems that we should be able to go beyond existing routines, that is – reveal quick flexibility – even during non-conscious goal pursuit. (Hassin, Aarts, Eitam, Custers, Kleiman, 2009: 6)

La natura dinamica del mondo e la scarsità delle risorse cognitive dimostrano che la flessibilità debba essere spiegata da meccanismi non coscienti anche quando mancano routine pronte. In che modo questo avviene? Proprio attraverso il ricorso a quegli schemi e a quegli “abiti” già immaginazzinati in precedenza. Le passate esperienze permetterebbero in questo modo di agire automaticamente davanti a situazioni simili ma non ancora esperite.

In un modello di questo tipo la coscienza appare avere un ruolo di secondo piano nella costruzione e nel reperimento delle informazioni dall’ambiente infatti:

Habits are the result of practice. They provide us with a well-learned set of skills and schemas that are often contingent on, and orchestrated by current

goals and can run and interact with our environment without us paying conscious attention or forming an explicit intention or plan to perform the behavior. Accordingly, habits are commonly accompanied by decreased awareness of the environmental events and action components involved, which suggests that we delegate habits to the unconscious. Such habits may be difficult to overcome by an act of conscious will. Furthermore, although many of our habits can occur without conscious intervention, people sometimes reflect on their habits. (Aarts et.al.,2009: 326)

Ciò che stanno dicendo i sostenitori di questo modello è che a suffragare l'ipotesi che processi coscienti non giochino il ruolo di primo piano che si è soliti attribuirgli, sopraggiunge l'ulteriore constatazione che i processi avvengono senza consapevolezza e anche qualora questa vi fosse, spesso l'azione automatica è relativamente impermeabile e non modificabile. Tutto ciò sarebbe impossibile senza implicare una forma di flessibilità cognitiva che permette a tutte queste "situazioni" (o rappresentazioni di situazioni) di attivarsi a seconda delle esigenze imposte dal contesto. Ma in che modo è possibile che schemi immagazzinati in precedenza possano essere riutilizzati in assenza di routine? La questione in realtà è molto complessa ed è bene sottolineare che questi approcci non rappresentano una soluzione al problema ma piuttosto una descrizione di quest'ultimo. Rimane infatti inspiegato il modo in cui concretamente riusciamo ad immagazzinare le informazioni dai differenti contesti nei quali si presentano, e soprattutto il modo in cui riusciamo ad utilizzare queste routine e questi schemi già appresi precedentemente.

2.4 Flessibilità cognitiva e meccanismi gerarchico-associativi

Ad integrazione degli approcci basati sulla pura automaticità, si colloca la posizione di John Duncan che può essere considerato uno dei più influenti sostenitori del modello cognitivo gerarchico associativo. Il modello da lui proposto, il *multiple demand system* (MD) fornisce una possibile risposta al problema dell'acquisizione e della gestione di nuove informazioni:

In all realistic behaviour, from cooking a meal to solving a problem in formal logic, goals are achieved by assembling a series of sub-tasks, each separately defined and solved. In this respect, human behaviour resembles the sequential activity of conventional computer programs, assembling a series of operations that together achieves the final goal. The problem is well illustrated by artificial intelligence systems, from the earliest days of the General Problem Solver to systems such as Newell's SOAR and Anderson's ACT. In all these systems, as in all complex human behaviour, the problem is not simply to control isolated steps of thought or behaviour, but equally significant, a task must be decomposed into useful parts, with these component sub-tasks addressed in turn. It has long been noted that, after major frontal lobe damage, the structure of complex behaviour is impaired, with important parts omitted, irrelevant parts introduced, and the whole failing to achieve its ends. Such results indicate an obvious link to the problems of symbolic artificial intelligence, and to the sequential programming of goal-directed behaviour. Sequential mental programming has several requirements. Each step requires focus, that is enclosed processing only of the parts of the current sub-task. As sub-tasks are completed, the current cognitive focus must be abandoned and a new focus created. Often, selected results from one sub-task must be passed to the

next. Especially in novel tasks, useful subparts must be defined and separated. Mental programming of this sort is an indispensable part of all structured mental activity. I suggest that neural activity in MD [multiple-demand] cortex¹⁰ is well suited to such programming, explaining broad recruitment across cognitive domains, and a central role in human intelligence. (Duncan, 2010: 172)

Ciò che propone è un modello fondato su un'organizzazione gerarchica dell'informazione, in cui la working memory ha un ruolo fondamentale. Si tratta di un sistema che prevede l'impacchettamento dell'informazione in sotto-componenti via via più semplificate.

La *ratio* alla base del suo modello è rappresentata dalla convinzione che le informazioni necessarie per molti compiti cognitivi siano troppo complesse per essere processate dalla memoria di lavoro. Di conseguenza, affinché quest'ultima lavori correttamente, è necessario ridurre le informazioni in sotto informazioni via via meno complesse, processabili on line ed in maniera veloce ed efficace. La complessità della realtà (delle informazioni) richiede l'intervento di una struttura complessa e gerarchica necessaria a semplificarla. Questo perché secondo Duncan, a compiti complessi segue una minore probabilità di scegliere l'informazione pertinente e dunque una minore probabilità di arrivare a perseguire lo scopo che ci si era prefissati. I compiti complessi devono dunque essere convertiti in modo da permettere ai processi attentivi e alla working memory di elaborarli. È proprio sulla nozione di working memory che Duncan si

¹⁰ In questo caso Duncan con " MD cortex " vuole indicare quelle regioni cerebrali in cui opera il sistema da lui ipotizzato: " This multiple-demand (MD) pattern extends over a specific set of regions in prefrontal and parietal cortex, in particular: cortex in and around the posterior part of the inferior frontal sulcus (IFS), in the anterior insula and adjacent frontal operculum (AI/FO), in the pre-supplementary motor area and adjacent dorsal anterior cingulate (preSMA/ACC), and in and around the intraparietal sulcus (IPS)"

concentra, mettendola in discussione e modificandone il significato. A tal proposito introduce quello che definisce il *task model* Duncan et.al. (2008):

Typically, this model includes both the current state of relevant aspects of the world, goal states to be achieved, and candidate methods for achieving them. The contents of this model represent the active body of knowledge that bears on current decision making[...]. As a new task is attempted or new instructions are received, a new task model must be constructed. This model should incorporate all the knowledge that bears on current activity, organized into suitable chunks or subroutines of a mental program. These chunks must somehow be established in a form that can later be used when suitable trigger conditions occur. (Duncan et.al.,2008: 140)

In cosa il modello ipotizzato da Duncan, dunque la sua concezione di working memory, differisce dalla sua tradizionale versione? Vediamo in che modo lo studioso chiarisce i termini di questa differenza:

Evidently, the task model is a form of working memory. How does it relate to other kinds of working memory limit? One popular conception, for example, combines a central executive with several lower level slave systems, each storing different kinds of information (Baddeley, 1986). Others equate working memory with immediate awareness for a few chunks of information (Cowan, 2000). In several respects, models such as these address processing limits that seem rather different from the one we are considering here. One obvious difference concerns capacity. Traditional working memory experiments focus on the ability to remember just a few stimulus items, for example, typical spans of six or seven digits or five or six

spatial locations. Correspondingly, traditional models of working memory have a limit of a few chunks of information[...]. It is often proposed that the few chunks of working memory are closely related to the immediate contents of awareness. In contrast, nobody would suppose that a whole set of instructions is kept immediately in awareness until the time the instructions are used. A second difference concerns dynamics. Traditional forms of working memory are highly labile—for example, emptying when a string of digits is reported and filling again with the digits for the next trial. In contrast, the task model must be a relatively stable control structure, surviving many intervening events to control performance when appropriate trigger conditions arise. (Duncan et.al.,2008: 140)

Il primo aspetto che va sottolineato al fine di comprendere le differenze tra la sua concezione di working memory e quella tradizionale riguarda i “limiti” della capacità di processamento della working memory. La posizione di Duncan differisce infatti a tal proposito dalle più popolari presenti in letteratura. Lo studioso paragona la working memory alla “long term memory” utilizzata da Ericsson e Kintsch (1995). Secondo questi ultimi:

storage in LTM¹¹ is assumed to be primarily associative, relating different items to one another and relating items to attributes of the current situation (current context). The time required for storage of a new retrievable memory trace in LTM has been estimated to be relatively long--about ten seconds (Simon, 1973).

¹¹ Long term memory

Dall'adozione di questa concezione di memoria ne deriva la differenza tra il tradizionale modo di concepire la *working memory* e il modello ipotizzato da Duncan che, come lui stesso sottolinea, si rivela palese se si considerano due aspetti: la capacità (*capacity*) di "tenere in mente" informazioni, e la dinamica (*dynamics*) in cui tale processamento/immagazzinamento avviene.

A proposito della prima caratteristica (*capacity*) la tradizionale concezione della *working memory* la configurerebbe come capace di contenere solo una limitata quantità di informazioni per breve tempo ("few chunk of informations") e per di più relative alla situazione presente. Nel *task model* (e nell'idea di *working memory* di Duncan) invece, in accordo con l'idea di memoria a lungo termine, il tempo di permanenza delle informazioni è più lungo e permette un richiamo alla mente di situazioni anche lontane nello spazio e nel tempo.

Per quanto riguarda la dinamica (*dynamics*), invece, la differenza si traduce nel fatto che la *working memory* tradizionale è labile, si riempie e si svuota velocemente ed è condizionata dal contesto presente. La *working memory* così come ipotizzata da Duncan, invece, servendosi di informazioni non per forza immediatamente derivabili dal contesto presente, deve garantire stabilità e perdurare maggiormente nel tempo.

Ciò che si evince dall'ipotesi di Duncan, è che questi meccanismi permettono all'uomo di agire e modificare l'ambiente in cui vive in maniera creativa, scegliendo la situazione che gli sembra più pertinente.

Potremmo dunque dire che il comportamento umano, si distingue da quello delle altre specie per la flessibilità rispetto al contesto. Duncan introduce a questo scopo una comparazione tra l'intelligenza animale e quella umana per sottolinearne le differenze. A tal proposito introduce il concetto di IRM "innate releasing mechanism" di Lorenz:

The IRM is conceived as a neural process delivering a fixed fragment of behavior when released by a suitable triggering event—the moth flashing opens its wings to reveal staring eye spots when touched by a predator, the toad orienting to the sight of a worm, or the human urged to protect an infant by the sight of large eyes and tall forehead. In much animal behavior, concatenation of such behavioral fragments, each controlled by its own IRM, produces complex, goal-directed sequences or programs of activity; for example, when two mating sticklebacks, each responding to the actions of the other, proceed by a series of stages into the nest where spawning takes place (Tinbergen, 1951). In human cognition, too, thought and behavior unfold in a complex, structured sequence, with many component fragments assembled to achieve short- and long-term goals (Miller et al.,1960). By comparison with mating sticklebacks, however, human thought and behavior have essentially infinite flexibility and complexity. Unlike the IRM, fragments of human cognition must be momentarily constructed, shaped by the arbitrary requirements of current activity. (Duncan,2013: 35)

Ciò che sta dicendo Duncan è dunque che il comportamento animale è rigido e in larga misura obbligato e limitato dalle strutture stesse che lo rendono possibile. In questo modo non c'è alcun bisogno di impacchettare in maniera semplificata la realtà perché il meccanismo di IRM da solo basta a controllare il comportamento animale. Ciò non significa che da questa rigidità non possano nascere comportamenti complessi e diretti ad uno scopo. Semplicemente quello scopo è preordinato perché inserito nella storia della loro specie. La peculiarità del comportamento umano, al contrario, si traduce nella flessibilità delle strutture che lo rendono possibile. Questo ci permette di manipolare l'ambiente

e le situazioni.¹² Ma come fare a gestire la quantità di informazioni che riceviamo dall'ambiente? Ciò ci conduce all'inevitabile bisogno di semplificare:

human thought and behavior have essentially infinite flexibility and complexity. Unlike the IRM, fragments of human cognition must be momentarily constructed, shaped by the arbitrary requirements of current activity. [...] The essential difficulty is that, if all aspects of the problem are considered at once, the search space of possible alternative routes is simply too large and unconstrained, resulting in many suboptimal choices. In the second solution, the architecture tends first to choose a relatively abstract subgoal and then work within that to elaborate a detailed solution. Now the path to the goal is orderly and direct, as one organizing subgoal after another comes into force and controls system function until it is achieved. (Duncan,2013: 35)

L'approccio di Duncan potrebbe essere inquadrato all'interno di un altro approccio, quello probabilistico:

Probability theory forms a natural framework for explaining the impressive success of people at solving many difficult inductive problems, such as learning words and categories, inferring the relevant features of objects, and identifying functional relationships. Probabilistic models of cognition use

¹² Duncan sta affermando, con altre parole, ciò che molta antropologia filosofica afferma da tempo. L'idea è che la specie umana si distingua dalle altre specie per la sua "incompletezza" (riferendosi agli istinti). Già Herder (1770) e Gehlen(1940), e più tardi Leroi-Gourhan, hanno affermato l'idea che la grandezza dell'uomo stia proprio nella carenza dei suoi istinti. Citando Virno (2002: 26): "L'animale umano è più povero degli altri animali. Tale povertà consiste anzitutto in certi primitivismi organici e in un deficit di istinti specializzati. Un istinto specializzato consente di sapere a priori, con assoluta sicurezza, che cosa fare in questa o quella occasione determinata, per esempio quando si tratta di schivare un pericolo o di procacciarsi il cibo. L'uomo ha istinti deboli, generici: non sa con precisione che cosa fare, come comportarsi"

Bayes's rule to identify probable structures or representations that could have generated a set of observations, whether the observations are sensory input or the output of other psychological processes (Austerweil et.al.,2015: 187)

Ciò che li accomuna è, per rimanere su un piano generale, la convinzione che strutture complesse siano il frutto dell'interazione di processi più semplici. L'idea dunque che l'apprendimento avvenga attraverso un sistema gerarchico.

Un genere di approccio probabilistico è quello Bayesiano, in cui tra un ventaglio di ipotesi possibili disposte su livelli multipli (ordinati, appunto, gerarchicamente), viene selezionata la più probabile. Quest'ultima sarà poi confermata se ritenuta ancora la più probabile dopo l'osservazione dei dati provenienti dall'esperienza:

A probabilistic model starts with a formal characterization of an inductive problem, specifying the hypotheses under consideration, the relation between these hypotheses and observable data, and the prior probability of each hypothesis [...]. Probabilistic models therefore provide a transparent account of the assumptions that allow a problem to be solved and make it easy to explore the consequences of different assumptions. Hypotheses can take any form, from weights in a neural network to structured symbolic representations, as long as they specify a probability distribution over observable data. Likewise, different inductive biases can be captured by assuming different prior distributions over hypotheses. The approach makes no a priori commitment to any class of representations or inductive biases, but provides a framework for evaluating different proposals (Griffith et. al,2010: 358)

L'efficacia di questi meccanismi è data dalla capacità di coniugare vincoli e flessibilità. Quest'ultima è ottenuta, infatti, tramite lo scambio informativo su molteplici livelli, passando da meccanismi di basso livello a meccanismi di livello più alto:

Although this HBM imposes strong and valuable constraints on the hypothesis space of causal networks, it is also extremely flexible (Tenenbaum,2011: 1284)

In particolare il modello gerarchico Bayesiano permette di creare uno spazio di lavoro in cui le conoscenze a livello bottom possono essere dunque unite a quelle a livello top, in un processo di continuo scambio e accumulo di informazioni:

Hierarchical Bayesian models (HBMs) address the origins of hypothesis spaces and priors by positing not just a single level of hypotheses to explain the data but multiple levels: hypothesis spaces of hypothesis spaces, with priors on priors. Each level of a HBM generates a probability distribution on variables at the level below. Bayesian inference across all levels allows hypotheses and priors needed for a specific learning task to themselves be learned at larger or longer time scales, at the same time as they constrain lower-level learning. (Tenebaum,2011: 1282)

Abbiamo finora ripercorso ed analizzato le principali posizioni presenti in letteratura sulla flessibilità cognitiva, abbiamo in particolare analizzato la

posizione di Carruthers e di Sperber ed il loro tentativo di coniugare modularità e flessibilità cognitiva. Abbiamo visto che per coniugare queste ultime è necessario ipotizzare un sistema, il *practical reasoning system* che insieme al modulo della produzione linguistica permette l'integrazione dei contenuti e rende possibile la comunicazione tra moduli. Abbiamo scardinato, attraverso le posizioni di Hassin, Bargh e Zimmerman, l'idea che flessibilità e automatismi non siano conciliabili, sottolineando i punti di forza di questa posizione e il tentativo, anche se non risolutivo, di spiegare in che modo è possibile da una situazione nota rendere conto di una situazione non nota. Abbiamo inoltre analizzato nel dettaglio le posizioni che vedono la flessibilità dipendente da meccanismi di tipo gerarchico-associativo sottolineando come questi approcci, come quello di Tenenbaum o di Duncan, forniscano invece una possibile e valida soluzione al problema sollevato da Hassin, Bargh e Zimmerman.

2.5 Flessibilità cognitiva e mental file

Un discorso a parte merita la posizione assunta recentemente da Kovacs (2015). Quest'ultima dopo aver analizzato la letteratura sull'*early mindreading* e dopo aver dunque constatato la precocità con cui si manifestano nel bambino abilità di teoria della mente, prova a fornire un'alternativa alle ipotesi finora formulate. La sua idea è che le nostre abilità di *mindreading* siano possibili in maniera precoce grazie alla capacità di formare e trattenere in maniera flessibile, in quelli che lei definisce *belief files*, rappresentazioni mentali sulle credenze altrui:

I will be proposing here that representations of others' beliefs are stored in special representational structures– 'belief files'- that are sustained in parallel with one's own representations of the real world. The representational skeleton of these belief files allows for functionally separate subcomponents that can be individually updated, and for a format that supports the rapid encoding and identification of belief related information (Kovacs,2015: 6)

In realtà la nozione di *belief file* è ben più antica ed è nota in filosofia e linguistica, per rendere conto di questioni sollevate da Russel e Frege. Ultimamente è stata utilizzata anche in un'area delle scienze cognitive, quella della psicologia dello sviluppo, con un significato differente e tuttora da approfondire dato il suo recente utilizzo¹³:

The concept of mental file reflects the realization that different concepts used in different areas of cognitive science share central features. Mental files play an important role in philosophy, addressing longstanding issues about Russell's problem of acquaintance and Frege's foundational problems of logics about identity and the sense-reference distinction. As discourse referents they play a role for solving problems of reference in linguistics. In psychology they have only been used as object files in research on attention and search and on tracking visually moving objects. They have been used in infancy research to explain infants' individuation of objects and their understanding of numerosity. They have not played any significant role in later child development. It has not yet entered research on children's theory of mind, which is surprising since mental files theory in philosophy has been

¹³ La questione è stata trattata recentemente nell'ultimo convegno EAP Cogsci 2015 (EuroAsianPacific joint conference on cognitive Science). In questo paragrafo viene presa in considerazione solo la posizione di Kovacs, perché affronta il tema della flessibilità. Un'altra posizione interessante sui mental file è quella di Perner (2015) che verrà considerata nel paragrafo 3.1.

used extensively to deal with the pernicious logical problems created by statements about beliefs and other mental terms. (Leahy,2015: 34)

Potremmo ridefinire i *belief file* come una sorta di *dossier* che contengono rappresentazioni mentali della nostra conoscenza sul mondo. Ciò su cui si concentra Kovacs (2015), è ciò che lei definisce *empty belief file*. In sostanza si chiede cosa succede quando non possediamo che una rappresentazione parziale dell'inferenza che dobbiamo predire:

The centerpiece of this exploration was the concept of an empty belief file, i.e., a structure that allows for the content component to be tagged by a placeholder and, importantly, to fill in its content later in time. This flexibility may be crucial for a developing cognitive system that does not have the resources to encode all possible information at once. But even a mature system can make good use of such a feature, as there are many situations where the content of another agent's belief is not directly available. However, if we can encode that there is an agent who has a belief, although we don't know exactly what this belief might be, we can fill in this content information later, when it becomes available. Importantly, it is not clear how holding or updating a belief without knowing its content can be theoretically possible, unless one argues for a framework where attributing a belief and computing its content are separate components. (Kovacs, 2015: 16)

Quello che sta supponendo Kovacs (2015) è che alla base della possibilità di fare inferenze sugli stati mentali ci sia una struttura rappresentazionale capace di esercitare la sua "predizione" sul comportamento altrui anche quando quest'ultima è priva di contenuti completi e specifici. Non solo. Sta

anche dicendo che la possibilità di utilizzare in maniera efficace questa stessa struttura, anche in assenza di contenuti specifici, è condizione necessaria affinché si manifestino precocemente le abilità di *mindreading*. La flessibilità è decisiva in un sistema di questo tipo perché permette di codificare informazioni che prevedono per loro natura un agente e una credenza, anche se questi ultimi non sono ancora definiti e saranno disponibili solo più tardi. È dalla flessibilità di queste strutture che è possibile infatti in pochi secondi adattare (nonostante la parzialità delle informazioni) il nostro comportamento sulla base della rappresentazione dell'altro (che nel frattempo abbiamo creato nella nostra mente):

While we might encounter such situations, and even more complex ones, on a daily basis, note that updating an attributed belief multiple times involves a set of rather complex processes. In the span of a few seconds, we actually have to implement a large number of computations regarding what the other person believes, and about how she would behave on the basis of those beliefs. Furthermore, we use the outcome of these computations to flexibly adapt our own behavior as a function of the anticipated behavior of the other person. (Kovacs,2015: 5)

D'altronde sarebbe impossibile creare dal nulla rappresentazioni complesse senza sistemi che permettono di accumulare conoscenze seppur parziali sul mondo. L'esempio che Kovacs (2015) fornisce a tal proposito è questo: Siamo impegnati in una conversazione con una collega e ad un certo punto vediamo che un carrello sta scivolando verso di lei. Non sappiamo se lei ha visto oppure no quel carrello perché, se è vero che sembrerebbe essersi girata verso quest'ultimo, di fatto sembra non interessarsi alla questione, magari perché impegnata a difendere la sua tesi nel discorso che sta intrattenendo, magari invece perché pur essendosi girata non si è accorta

del carrello. Come facciamo ad elaborare tutte queste informazioni in così poco tempo? In nessun altro modo se non presupponendo che grazie ad una forma di flessibilità cognitiva che caratterizza le strutture in questione, è possibile da una piccola informazione accumularne delle altre in un file che si riempie e che si aggiorna continuamente fino al punto da permettere una percezione *spontanea* del comportamento e delle intenzioni dell'altro coinvolto nell' interazione:

It is quite unlikely that we only infer our colleague's belief at the very last moment, right before she would get hit by the cart. Indeed, this would leave little time to update our representation of her belief multiple times. Presumably, in such situations we track others' beliefs spontaneously and update them in a continuous manner (Kovacs, 2015: 5)

E ancora continua:

Successful ToM reasoning requires keeping track of other agents' mental states (e.g., their beliefs), which can be realized through opening a belief file that can then be filled with various contents. Belief files provide a representational structure with variables for (1) the agent, as the belief holder and for (2) the belief-content, in a way that each can be separately updated (Kovacs,2015: 7)

La constatazione che ha portato Kovacs (2015) a ipotizzare un sistema predittivo di questo tipo, è prima di tutto e soprattutto teorica. In sostanza ciò che si è chiesta è: che cosa succederebbe se ci trovassimo davanti ad una

situazione astratta al punto da non sapere chi compie l'azione ma da sapere che ad un certo punto ci sarà qualcuno che agirà in qualche modo? Probabilmente non cancelleremmo il fatto noto e cioè il fatto che *qualcuno compirà un'azione*. Questa è la condizione affinché un *belief file* sia aperto:

At a minimum, online belief formation requires a potential agent; if there is no agent, no belief file can be opened. (Kovacs,2015: 17)

Kovacs (2015) sottolinea inoltre che queste abilità sono possibili attraverso una co-occorrenza di fattori che regolano il funzionamento processuale del sistema. Da qui ne deriva che 1. l'abilità di mindreading non può essere considerata un'abilità "tutto o nulla"¹⁴. Quest'ultima è piuttosto un processo in divenire costruito storicamente. 2. i fattori che possono causare deficit nella rappresentazione delle intenzioni/comportamenti altrui possono essere molteplici:

Thus, solving a typical false belief problem involves a larger set of cognitive processes, such as: i) opening a belief file; ii) computing the content of someone's belief; iii) binding belief representations to corresponding agents; iv) performing belief evaluations and pondering between true and false beliefs, and v) making implicit or explicit inferences upon false beliefs for behavioral predictions. [...] Therefore, if an individual shows difficulties in reasoning about others' mental states on such tasks, this may result from problems in conceiving that agents have mental states and in opening a belief file; from problems in computing the content of such a belief; from a failure in dealing with the conflict between different belief representations

¹⁴ In questi caso ci si riferisce al suo sviluppo ontogenetico, anche se la constatazione vale ed è riferibile anche a livello dello sviluppo filogenetico.

(one's own true belief and the attributed false belief), or from difficulties in making behavioral predictions (Kovacs,2015: 6-7)

Nulla da eccepire alla coerenza teorica e alla capacità esplicativa di questo modello, molti dubbi rimangono però sulla concretezza e soprattutto sulla sua reale spendibilità in contesti differenti da quelli esclusivamente speculativi. Non è chiaro, infatti, in cosa si traduca a livello neurobiologico, un *belief file*, e a maggior ragione un *empty belief file*. L'argomentazione di Kovacs (2015) tra l'altro, sembra possedere molte delle caratteristiche dei modelli probabilistici con cui il suo modello condivide gli aspetti positivi: l'accumulo con conseguente immagazzinamento dell'informazione, l'organizzazione gerarchica dei dati percepiti (necessaria in un modello come il suo, e più in generale nei modelli che prevedono un accumulo delle informazioni), l'idea che l'apprendimento sia possibile attraverso una forma di predizione del non noto attraverso il non (o poco) noto.

2.6 Flessibilità, associazioni e generalizzazioni

Abbiamo finora analizzato in che modo la letteratura si occupa della questione della flessibilità cognitiva. L'abbiamo fatto con lo scopo di arrivare a comprendere in che modo la flessibilità entra in gioco nella costruzione delle abilità di mindreading.

Andiamo adesso ad analizzare degli altri modelli cognitivi e andiamo a vedere in che modo essi possono aiutarci a ipotizzare quel modello cognitivo fondato su una nuova e differente idea di associazione, che avevamo annunciato all'inizio e che può aiutarci a far luce sul modo in cui riusciamo a percepire le intenzioni altrui. Andiamo ad analizzare in particolare in che

modo è possibile rendere conto della flessibilità mediante meccanismi associativi.

Ma che tipo di associazione? Se analizziamo l'idea classica di associazione non è chiaro in che modo la flessibilità giocherebbe un ruolo fondamentale. Il modello associativo fondato sul classico schema stimolo-risposta non consente di spiegare la flessibilità cognitiva.

É in realtà ormai abbastanza condivisa l'idea che parlare di associazione non significhi solo parlare di rigidi legami tra input e output. Come fa notare Smith (2000) probabilmente la diffidenza, da parte delle scienze cognitive, nei confronti dei modelli associativi deriva dall'erroneo accostamento di questi ultimi con i modelli comportamentisti. In realtà il comportamentismo non apprezza l'associazionismo perché quest'ultimo sottolinea l'importanza dei contenuti interni alla mente:

[...] radical behaviourism hated the idea of associationism. Why? Because associative learning is about ideas, about internal mental events, about the processes that make cognition (Smith,2000: 169)

É necessario dunque fare chiarezza e riabilitare la concezione che si ha degli approcci associativi.

La ricognizione delle posizioni sull'early mindreading ha sottolineato la debolezza delle argomentazioni fondate sull'innatismo e sulle regole comportamentali. Di fatto gli approcci associazionisti possono far luce e fornire nuove alternative nella spiegazione dell'apprendimento. Secondo Rakison e Lupyan (2008) che si sono occupati di comprendere in che modo i concetti si sviluppano nella mente del bambino, è impossibile pensare ad una visione della cognizione fondata sulla "core knowledge", cioè sull'idea che sia un corpo di conoscenze presente fin dalla nascita a garantire lo

sviluppo concettuale nel bambino. Ne deriva la loro critica nei confronti dell'architettura cognitiva su cui questi modelli si basano (modulare e specifica per dominio). Analizziamo le critiche nei confronti di quei modelli e proviamo a comprendere che tipo di alternativa forniscono i due studiosi. La loro posizione come spesso succede tra i difensori dell'apprendimento associativo, si basa sull'idea che il mondo sia retto da sufficienti regolarità statistiche, tali da permettere agli individui di creare associazioni e legami di causa-effetto tra di esse¹⁵. La seconda constatazione che li porta a criticare gli approcci fondati sull'innatismo è che secondo loro questi approcci non sono supportati da alcuna prova empirica. In particolare fondano la loro argomentazione su tre punti:

First, there is no direct empirical evidence that infants possess innate modules, innate core knowledge, or specialized mechanisms that allow them to encode distinct kinds of information (and in particular, motion characteristics). Indeed, the very notion of an innate specialized module or a core principal is difficult, if not impossible, to test empirically. Two lines of reasoning that are commonly adopted by core knowledge theorists are that if an ability or knowledge is present early in life it must be innate and that if infants find certain expectations about, for example, continuity difficult to learn then these expectations may be innate. Both of these assumptions logically are flawed: if infants display knowledge about an event at 3 months of age, they have had 3 months to learn about such events; if infants find it difficult to form an expectation for an event it does not necessarily mean that it is not learned. Second, the notion that infants possess two separate mechanisms for concept formation one for perceptual information and one for conceptual information (Mandler, 1992) has been labeled as

¹⁵ In particolare i due studiosi si riferiscono all'apprendimento dei concetti di animato/inanimato.

unparsimonious because it creates a heavy biological burden (Madole & Oakes, 1999; Quinn & Eimas, 2000; Quinn, Johnson, Mareschal, Rakison, & Younger, 2000; Rakison & Hahn, 2004). Third, although we agree that motion cues can be misleading (Gelman et al. 1995), there may well be sufficient regularities in the world for an associative mechanism to account for learning in this domain. (Rakison & Lupyan, 2008: 8)

Nella argomentazione dei due autori ancora più interessante della pars destruens è la pars construens, in cui si impegnano a fornire un modello della cognizione alternativo a quello in precedenza criticato. Definiscono questo modello *Constrained attentional associative learning* (CAAL) e lo descrivono come segue:

1. The primary mechanism that supports early concept formation is domain-general associative learning.
2. Conceptual development is best depicted as a continuous incremental augmentation of initial representations.
3. Infants possess a number of inherent perceptual biases that direct attention to specific aspects of the array.
4. There exist sufficient statistical regularities regarding static and dynamic cues within the category of animates and the category of inanimates.
5. The development of object concepts is explained by advances in information-processing abilities such as improving short- and long-term memory and increasing encoding speed, as well as neurological maturation.
6. Constraints on learning emerge as a product of prior experience with statistical regularities. (Rakison e Lupyan, 2008: 9)

In sostanza teorizzano un modello fondato interamente sull'apprendimento, che non ha dunque bisogno di ricorrere all'innatismo delle strutture. Forniscono inoltre una possibile soluzione al problema di

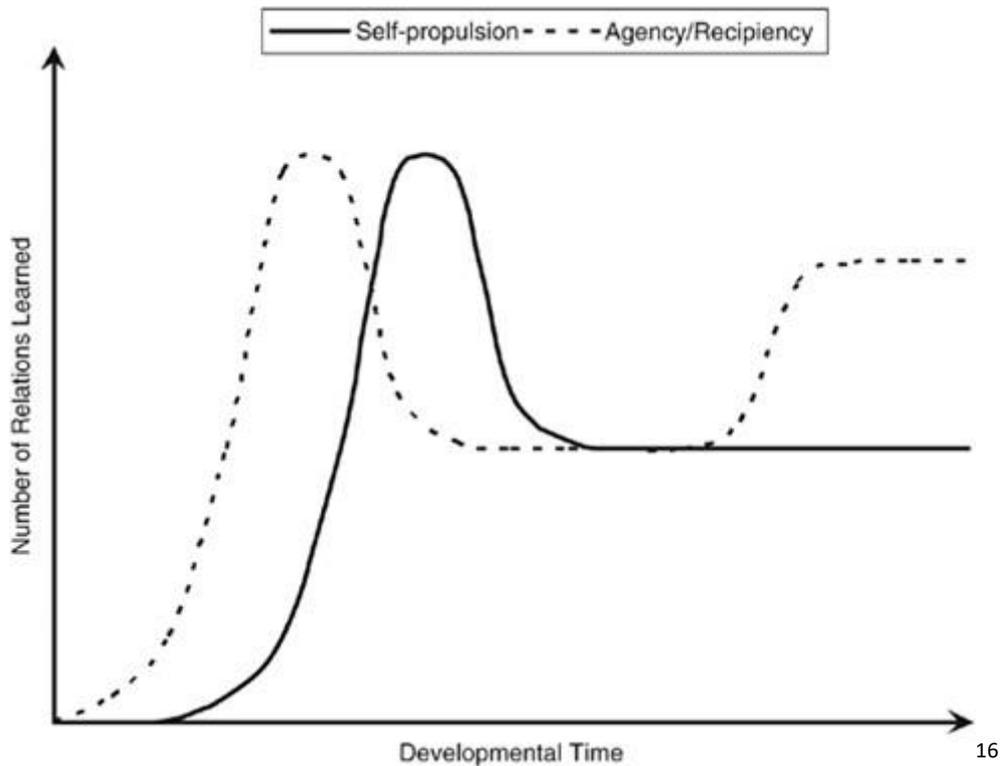
delimitare l'infinità di dati presenti nell'esperienza, attraverso la constatazione che le regolarità statistiche permettono di operare una scrematura sulla realtà percettiva, e fondano sul miglioramento delle abilità di memoria a breve e lungo termine lo sviluppo e l'accumulo della conoscenza concettuale:

[...] concept development is a process of continuous representational enrichment grounded in a perceptual system that is sufficiently sensitive to allow infants to form categories that cohere because of perceptible similarity relations. (Rakison e Lupyan,2008: 13)

La precocità di alcune abilità presenti nel bambino in questa visione è spiegate dunque, non attraverso un cambiamento qualitativo delle strutture che ne stanno alla base, ma piuttosto attraverso un cambiamento quantitativo di queste ultime:

[...] we propose that the basic mechanisms involved in concept development including associative learning do not change over time. Instead, our view is that infants are able to encode ever more complex information over time as a result of, initially, advances in information-processing abilities and an increasingly sensitive perceptual system and then, later, the emergence of linguistic skills. (Rakison e Lupyan,2008: 23).

Questo avanzamento nell'apprendimento è ben rappresentato nello studio di funzione che descrive la progressione del numero delle relazioni apprese (nell'asse delle ordinate) nello sviluppo temporale (nell'asse delle ascisse):



16

Una questione importante che è in parte affrontata anche da Rakison e Lupyan (2008) è quella delle modalità attraverso le quali è possibile l'astrazione di informazioni generali da contesti specifici. In particolare per affiancare una visione alternativa a quella innatista è necessario risolvere il problema che si deve affrontare se si fa derivare la conoscenza precoce da strutture apprese e non innate: da dove deriva, se non è innato, l'insieme di conoscenze sul mondo che riusciamo precocemente ad utilizzare?

¹⁶ Rakison (2006). Make the first move: How infant learn about self-propelled objects, *Developmental Psychology*, 42, 900-912. Il grafico si riferisce in realtà al modo in cui si sviluppa la capacità di percepire il movimento degli oggetti, e al suo covariare nel tempo. In questo contesto verrà utilizzato, in maniera più generale, per mostrare in che modo l'accumulo di relazioni apprese cresce in funzione del tempo a supporto di una visione anti-innatista.

HBM [Hierarchical Bayesian models] may answer some questions about the origins of knowledge, but they still leave us wondering: How does it all start? Developmentalists have argued that not everything can be learned, that learning can only get off the ground with some innate stock of abstract concepts such as “agent,” “object,” and “cause” to provide the basic ontology for carving up experience. Surely some aspects of mental representation are innate, but without disputing this Bayesian modelers have recently argued that even the most abstract concepts may in principle be learned. For instance, an abstract concept of causality expressed as logical constraints on the structure of directed graphs can be learned from experience in a HBM that generalizes across the network structures of many specific causal systems. Following the “blessing of abstraction,” these constraints can be induced from only small samples of each network’s behavior and in turn enable more efficient causal learning for new systems. (Tenenbaum,2011: 1284)

Per Tenenbaum l’insieme delle conoscenze sul mondo deriva proprio dalla generalizzazione che abbiamo chiamato in causa prima.

Se si utilizza una visione fondata sulle reti Bayesiane, è possibile pensare che anche la struttura più semplice (che di solito si pensa sia innata), possa essere il frutto di un processo di apprendimento che avviene attraverso una generalizzazione di strutture specifiche.

Non è difficile ipotizzare che i modelli della cognizione analizzati fino ad ora possano essere anche considerati responsabili della nostra percezione degli stati mentali altrui. In questa visione gli stati mentali non sarebbero altro che generalizzazioni di inferenze provenienti dall’esperienza attraverso meccanismi associativi e organizzati gerarchicamente. Nei prossimi paragrafi analizzeremo nel dettaglio in che modo questi modelli cognitivi

possono essere utilizzati per costruire un nuovo modello cognitivo che possa spiegare il modo in cui riusciamo a percepire gli stati mentali altrui in maniera precoce.

2.7 Flessibilità cognitiva e teoria della mente

In che modo la flessibilità cognitiva entra in gioco nella comprensione delle intenzioni altrui? Nei paragrafi precedenti abbiamo visto come alcuni indirizzi della ricerca attuale spiegano il nostro modo di agire in maniera flessibile rispetto al contesto. In questo paragrafo proveremo a far luce sul legame che intercorre tra l'agire flessibile e le abilità di *mindreading*. In particolare proveremo a costruire un modello cognitivo che sia in grado di spiegare queste ultime senza chiamare in causa né l'innatismo né le regole comportamentali. A tal proposito, iniziamo fornendo una definizione di flessibilità che può essere utile ai nostri fini:

Cognitive flexibility is the ability to appropriately adjust one's behavior according to a changing environment. Cognitive flexibility enables an individual to work efficiently to disengage from a previous task, reconfigure a new response set, and implement this new response set to the task at hand (Dajani, Uddin, 2015: 1)

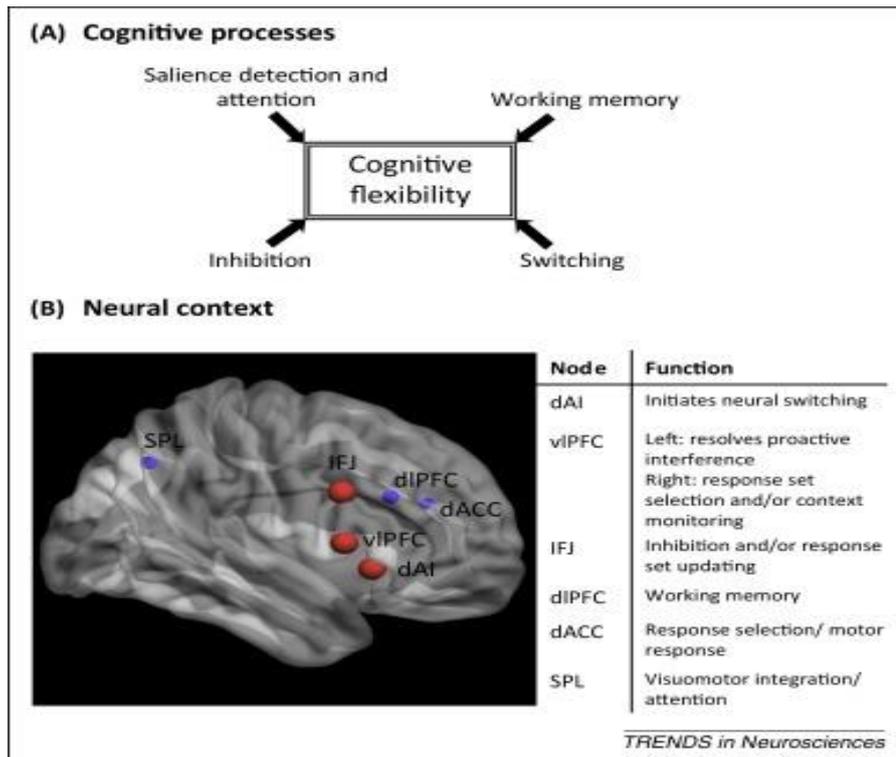
La capacità di sganciarsi da un compito per iniziarne un altro, riconfigurare una nuova situazione sulla base delle nuove informazioni, tenendo a mente lo scopo del compito che deve essere eseguito, è dunque una delle possibili definizioni di flessibilità cognitiva. Quest'ultima, definita in questo modo, sembra avere molto in comune con le abilità necessarie a superare compiti di teoria della mente. Abbiamo visto, ad ulteriore conferma di questa affermazione, che una delle difficoltà dei bambini nei compiti di falsa credenza sia proprio l'incapacità di comprendere che esiste una prospettiva, e dunque una credenza, differente dalla

propria. Che tipo di rapporto esiste dunque tra queste due abilità (flessibilità e teoria della mente)?

L'idea di base di questo lavoro è che il loro rapporto sia molto stretto, in particolare, che la flessibilità cognitiva abbia un ruolo cruciale nella costruzione di abilità di teoria della mente. Vediamo perché:

In constantly changing environments, individuals must first identify how their surroundings have changed by directing attention to those elements that are in flux. After ascertaining that a previous strategy is not appropriate in the new environment, individuals must inhibit previous responses and reconfigure a new strategy. Individuals take in information and manipulate it in real time to flexibly switch responses from one scenario to another. Cognitive flexibility is not merely the sum of implementing various EFs but also requires shifting, or the reconfiguration of one's response set to the new goal. (Dajani, Uddin,2015: 3)

È interessante notare come da questo studio (Dajani, Uddin,2015) emerga con chiarezza la funzione della flessibilità nei processi cognitivi. Quest'ultima è necessaria per far fronte agli stimoli provenienti dall'ambiente in costante cambiamento, inoltre supervisiona e rende possibile il passaggio (shifting) da una situazione ad un'altra e la riconfigurazione di quella stessa situazione sulla base del nuovo obiettivo da raggiungere.

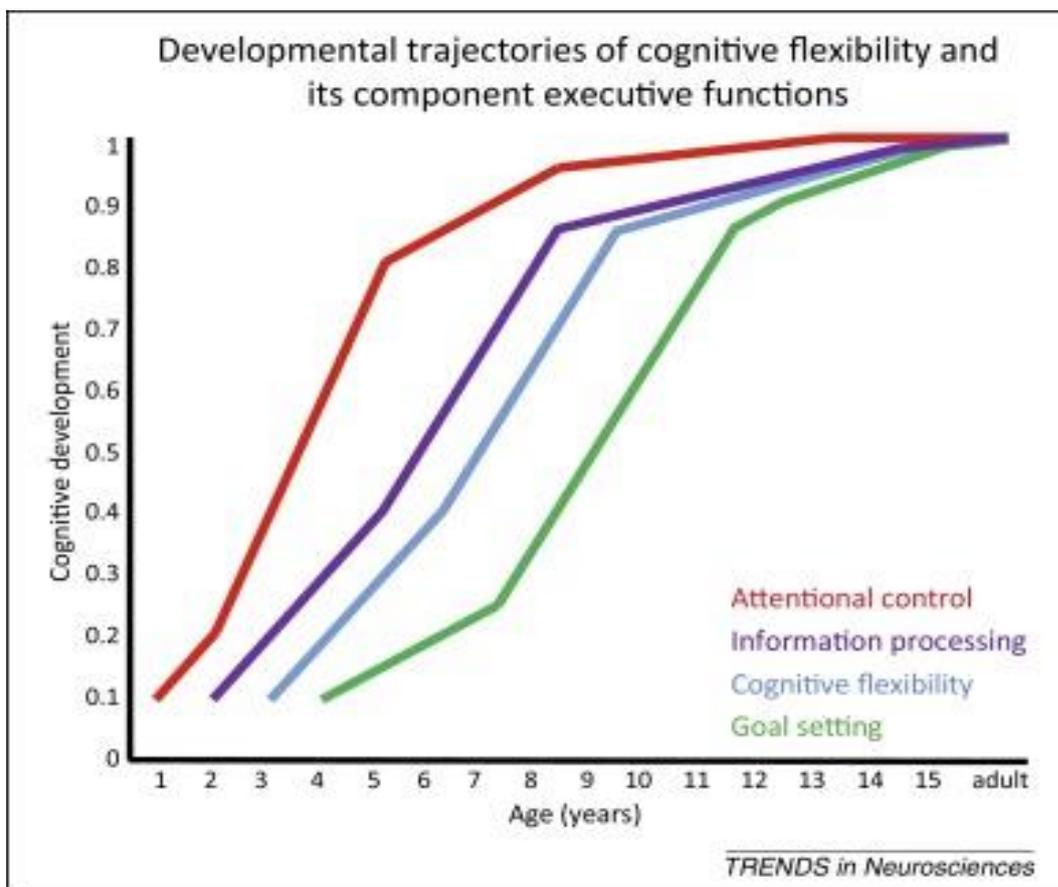


Dajani, Uddin, 2015

Riuscire a distinguere la flessibilità dalle differenti componenti delle funzioni esecutive in essa coinvolte è un compito molto difficile, ma è comunque possibile tracciare una traiettoria di sviluppo, in cui risulta chiaro che la flessibilità cognitiva si serve di altre abilità come il controllo inibitorio e la working memory:

Cognitive flexibility skills begin to develop in early childhood, with a sharp increase in abilities between 7 and 9 years of age. Cognitive flexibility becomes largely mature by 10 years of age but skills continue to improve throughout adolescence and into adulthood, reaching their peak between the ages of 21 and 30 years. Components of EF involved in cognitive flexibility follow different developmental trajectories. Inhibition develops as early as 12 months and is largely mature by 10–12 years of age. Working memory also emerges early in toddlerhood and continues to improve throughout adolescence. (Dajani, Uddin, 2015: 4)

In questo studio di funzione è possibile osservare la traiettoria di sviluppo delle componenti delle funzioni esecutive (attentional control, informative processing, cognitive flexibility, goal setting) in rapporto all'età dell'individuo:



Dajani, Uddin, 2015

Il rapporto tra flessibilità cognitiva e teoria della mente, è stato indagato anche da Kissine (2012), in un articolo in cui si preoccupa di chiarire in cosa si traduca concretamente il legame che intercorre tra abilità pragmatiche e teoria della

mente. Prendendo in esame il caso dell'autismo, viene messo in luce il ruolo che la flessibilità cognitiva, che permette al soggetto di cambiare prospettiva (o regola da seguire) in accordo con le esigenze imposte dalla situazione, potrebbe giocare nella costruzione di abilità di teoria della mente:

Envisaging reality under different aspects, that is, building and using alternative models of reality is what I will refer to under the term 'cognitive flexibility'. Although some inconsistencies subsist in the literature, difficulties in cognitive flexibility thus understood (not necessarily related to inhibitory control nor to working memory) emerge as being the executive dysfunction associated with ASDs (Hill, 2004; Ozonoff et al., 2005). It is also extremely plausible that failure on false-belief tasks can be partly explained, in the ASDs population, by a deficit at the level of cognitive flexibility (Kissine,2012: 8)

Kissine (2012) comparando le abilità di soggetti con sindrome dello spettro autistico e bambini di età inferiore ai quattro anni, constata come rendere saliente la prospettiva (credenza) del protagonista dell'azione favorisca la riuscita nei compiti di falsa credenza:

[...] Classical false-belief tasks involve certain cognitive demands that are distinct from belief attribution proper. Making more explicit the difference between the participant's own belief and the protagonist's belief—either by raising the prominence of the protagonist's belief or by rendering the participant's belief less salient—improves typically developing children's performance on false-belief tasks (although the developmental trend still

subsists, cf. Wellman et al., 2001). Rendering the mindreading target's perspective more salient can also improve the performance of people with ASDs on false-belief tasks. It is hard to decide whether facilitating classical false-belief tasks has the same effect on typically developing children and on people with ASDs, and, accordingly, whether both populations employ the same cognitive resources when they succeed on classical and/or modified false-belief tasks. For now, the important point is that false-belief tasks are made easier when an alternative perspective—that of the protagonist—is brought into focus. (Kissine, 2012: 5)

Se da un lato i soggetti affetti da sindrome dello spettro autistico sembrano mantenere intatte le abilità che permettono di cogliere gli aspetti salienti all'interno del contesto comunicativo al pari dei bambini al di sotto dei 4 anni, non è ancora chiaro il motivo per cui non riescano a cogliere il carattere intersoggettivo dello stimolo comunicativo e ad adottare dunque la prospettiva dell'interlocutore.

Per comprendere in che modo la flessibilità entra in gioco nella costruzione delle abilità di teoria della mente, dobbiamo però richiamare in causa le caratteristiche dei modelli probabilistico-associativi descritti nel precedente paragrafo, e introdurre la teoria della concettualizzazione situata di Barsalou (2005;2006;2016):

The construct of situated conceptualization developed originally to explain how simulations of conceptual knowledge become situated. The construct was also used initially to explain social embodiment effects, a form of social priming. Recently, situated conceptualization has been developed as a more

general theoretical framework that underlies perception, action, cognition, social cognition, affective processing, and appetitive processing (Barsalou,2016: 6)

In quest'ultimo e recentissimo articolo Barsalou (2016) riprende la teoria della natura situata dei concetti, precedente formulata, e la applica al *social priming*. Stando alla sua teoria, l'elaborazione e l'immagazzinamento dei concetti non possono prescindere dalle informazioni contestuali in esso contenute, in questo senso ogni concetto sarebbe appunto legato alla situazione (contestuale) in cui è stato esperito:

[...] the abstraction process stores extensive information about background situations while establishing concepts in memory. Later, during all forms of conceptual processing, these situations become active and influence performance. (Barsalou, Yeh, 2006: 351)

In particolare secondo Barsalou:

[...] even when people focus attention on a particular entity or event in perception, they continue to perceive the background situation; the situation does not disappear. If perceptual experience takes the form of a situation, and if a conceptual representation simulates perceptual experience, then the form of a conceptual representation should take the form of a perceived situation. When people construct a simulation to represent a category, they should tend to envision it in a relevant perceptual

situation, not in isolation. When people conceptualize chair, for example, they should attempt to simulate not only a chair but a more complete perceptual situation, including the surrounding space and any relevant agents, objects, and events. (Barsalou, Yeh,2006: 352)

Quest'idea che l'astrazione dei concetti porti con sé tutto il bagaglio di informazioni presenti nella situazione contestuale permette di pensare alla percezione come ad un fenomeno "attivo" nella costruzione del pensiero, in cui è possibile immagazzinare l'informazione percettiva e riutilizzarla di fronte a situazioni simili che ci appaiono già familiari.

Cosa succederebbe se provassimo ad applicare questa teoria al modo in cui riusciamo a percepire gli stati mentali altrui? Che tipo di architettura cognitiva è necessaria per permettere questa modalità di elaborazione dell'informazione?

L'idea di Barsalou (2016) è che l'informazione sia raggruppata, non attraverso l'astrazione di singoli concetti (o inferenze), quanto piuttosto attraverso l'impacchettamento di concetti (o inferenze) multimodali:

[...] the situated conceptualization framework offers a principled account of the knowledge structures that develop to produce these packages of multimodal inferences. Every time a situation is experienced, the states of all systems active at the time in the brain's situation processing architecture are captured in memory as a situated conceptualization. Rather than being stored as abstract symbolic propositions that describe these states, the actual states of these systems themselves are stored. On later occasions, when situational cues activate this situated conceptualization, they partially reactivate these states in a reconstructive manner, thereby implementing a multimodal simulation of the processing that occurred during the situation. From this perspective, social priming — and pattern completion inferences

in general — reflect the reinstatement of previous situational activity, guiding current situational activity in the moment. (Barsalou,2016: 9)

In modelli di questo tipo appare ovvio che l'organizzazione e l'immagazzinamento dell'informazione debba avvenire attraverso una rete di informazioni ordinate gerarchicamente tali da permettere uno scambio biunivoco tra processi a livello top e processi a livello bottom. Il modello che meglio rispecchia un'architettura cognitiva di questo tipo è rappresentato dalle reti bayesiane, come lo stesso Barsalou riconosce:

A natural way of thinking about the space of possible simulations within a simulator is as a space of Bayesian priors, with some simulations being more likely than others. Furthermore, the strength of a given prior can be assessed empirically (rather than simply assumed), reflecting well-established and readily measured factors, including how frequently and recently category instances have been experienced, how ideal or preferred they are, their similarity to other instances, and so forth. (Barsalou,2001: 191)

Il meccanismo attraverso il quale si arriva a compiere inferenze mentalistiche, dunque, è lo stesso che sottostà al nostro modo di comprendere e di agire nella realtà che ci circonda, scegliendo tra un ventaglio di ipotesi quella che ci sembra la più probabile per arrivare allo scopo che ci siamo prefissati di raggiungere. Utilizzando dunque la teoria della natura situata dei concetti, e l'approccio probabilistico bayesiano è possibile operare una vera e propria generalizzazione delle inferenze utilizzando i dati provenienti dall'esperienza, che sono stati precedentemente immagazzinati (Mazzone,2014).

Un modello fondato sull'idea che l'immagazzinamento della nostra conoscenza sul mondo avvenga in maniera situata all'interno di

un'architettura di tipo bayesiano, permette di risolvere la questione dell'attribuzione di credenze rispettando il criterio della parsimonia esplicativa.

In un modello di questo tipo c'è ancora spazio per la coscienza? A questa domanda proveremo a rispondere nel prossimo paragrafo.

2.8 Flessibilità cognitiva, coscienza e modularismo

Abbiamo visto precedentemente che una parte di letteratura considera la flessibilità dipendente da meccanismi di tipo cosciente. Secondo alcuni studi, dunque, la peculiarità dell'essere umano di agire in maniera flessibile è possibile attraverso un ingente dispendio di risorse cognitive. Abbiamo anche visto, però, che un'altra parte di letteratura considera la flessibilità dipendente da meccanismi di tipo automatico (Hassin, Bargh, Zimmerman), sfatando così l'idea che la coscienza abbia un ruolo determinante nei processi cognitivi.

Prima di continuare a delineare il ruolo della coscienza, è opportuno notare che quest'ultimo appare più o meno determinante a seconda del modello di architettura cognitiva che si decide di adottare.

In un'architettura cognitiva di tipo modulare, per esempio, è impossibile pensare di poter prescindere dai meccanismi coscienti. Le caratteristiche del modularismo, primo fra tutti l'impermeabilità tra moduli, rendono necessario ipotizzare che un meccanismo, necessariamente cosciente, sopperisca alla loro incomunicabilità. Lo stesso Carruthers, come abbiamo visto precedentemente, per risolvere i problemi derivanti dal modularismo ha bisogno di ipotizzare l'esistenza del *practical reasoning system*.

A quest'ultimo viene delegato il compito di integrare le informazioni provenienti da domini differenti. Carruthers non può dunque far altro che chiamare in causa la coscienza. Il legame tra architettura cognitiva di tipo modulare e meccanismi coscienti sembra essere molto stretto per le necessità imposte dal modularismo stesso, paradigma che per diverso tempo è stato dominante in ambito cognitivo. Cosa succederebbe però se ci affidassimo ad una architettura cognitiva differente?

2.9 Reti di bayes: è ancora utile la coscienza?

Se non vogliamo aderire all'idea che la mente sia retta da un'architettura di tipo modulare, una possibilità è affidarsi a quella che in letteratura viene definita l'ipotesi del cervello bayesiano (Bayesian brain hypothesis), quest'ultima postula che:

[...] evolutionary selection should have resulted in neural and cognitive processing principles that approximate a statistical optimum. This implies that the brain maintains and continually updates a generative (predictive) model of its sensory inputs, which allows for inference on hidden environmental states that are hierarchically organized and cause the sensory inputs that the agent experiences (Diaconescu, Mathys et.al ,2014: 4)

L'idea di base è che il cervello sia stato plasmato dall'evoluzione naturale per aggiornare e implementare continuamente gli input provenienti dall'esterno, in un sistema la cui peculiarità non è più l'impermeabilità tra tipi di conoscenza provenienti da domini differenti, ma piuttosto la continua interazione tra informazioni di diverso tipo. L'organizzazione gerarchica tipica delle reti bayesiane, in particolare, permette che le informazioni a livello top comunichino con quelle a livello bottom creando un meccanismo

di scambio biunivoco in cui gran parte del lavoro può essere svolto in maniera automatica. In questa cornice, che ruolo possono assumere i meccanismi coscienti?

Per chiarire il ruolo della coscienza, o almeno una delle possibili funzioni che può assumere in un'architettura di questo tipo, vorrei servirmi di un recentissimo studio di Dehaene, King (2016). I due studiosi provano a far luce sulle modalità in cui avviene nel nostro cervello il *parsing* dei compiti cognitivi, e cercano di comprendere cosa distingue a livello cerebrale la percezione conscia da quella inconscia:

Parsing a cognitive task into a sequence of successive operations is a central problem in cognitive neuroscience. A major advance is now possible thanks to the application of pattern classifiers to time-resolved recordings of brain activity [electro-encephalography (EEG), magneto-encephalography (MEG), or intracranial recordings]. The method determines precisely when a specific mental content becomes explicitly represented in brain activity. Most importantly, the ability of these pattern classifiers to generalize across time and experimental conditions sheds light on the temporal organization of information-processing stages. To illustrate these ideas, we show how the decoding of MEG and EEG recordings can be used to track the fate of conscious and unconscious brain processes during simple masking and auditory novelty tasks. (Dehaene, King, 2016: 85)

Da diverso tempo nel campo delle neuroscienze cognitive, si cerca di indagare in che modo il nostro cervello organizza e processa i dati provenienti dall'esterno, ma negli ultimi anni grazie a tecnologie innovative, come quelle dell'elettro-encefalografia (EEG) e della magneto-encefalografia (MEG), è stato possibile comprendere molto di più sui meccanismi che regolano il nostro cervello.

Dahaene, King (2016) in particolare, utilizzano queste tecnologie per ipotizzare l'intervallo di tempo in cui uno specifico contenuto mentale viene rappresentato esplicitamente durante l'attività cerebrale e sono riusciti a definire le rappresentazioni coscienti come "segnali del cervello che perdurano nel tempo, senza essere pienamente stabili perché possono essere improvvisamente rimpiazzati appena un nuovo oggetto mentale diventa il focus del pensiero cosciente":

We have hypothesized that conscious perception corresponds to the entry of information into a global neuronal workspace (GNW), based on distributed associative areas of the parietal, temporal and prefrontal cortices, that stabilizes information over time and broadcasts it to additional processing stages (Dehaene and Naccache 2001; Dehaene et al. 2003, 2006). Even if the incoming sensory information is very brief, the GNW transforms and stabilizes its representation for a period of a few hundreds of milliseconds, as long as is necessary to achieve the organism's current goals. Such a representation has been called "metastable" (Dehaene et al. 2003) by analogy with the physics of low-energy attractor states, where metastability is defined as "the phenomenon when a system spends an extended time in a configuration other than the system's state of least energy" (Wikipedia). Similarly, conscious representations are thought to rely on brain signals that persist for a long duration, yet without being fully stable because they can be suddenly replaced as soon as a new mental object becomes the focus of conscious thought. (Dehaene, King, 2016: 93)

L'ipotesi di Dehaene, King (2016) è che la percezione cosciente avvenga all'interno di quello che definiscono "global neuronal workspace" (GNW) che ha il compito di trasformare le informazioni percettive provenienti

dall'esterno in rappresentazioni mentali e di stabilizzare queste ultime per il periodo necessario al raggiungimento dello scopo da parte dell'individuo.

I due studiosi per definire le rappresentazioni coscienti utilizzano una definizione proveniente dalla fisica: quella di metastabilità.

In fisica il termine metastabilità indica una particolare condizione di cui gode un sistema che si trova in equilibrio stabile fino a quando una fonte di energia non interviene a turbarlo. Allo stesso modo le rappresentazioni coscienti sarebbero metastabili in quanto rimangono stabili e perdurano nel tempo solo fino a quando un nuovo oggetto mentale non ne turba l'equilibrio.

La coscienza sembra dunque avere quel ruolo che hanno le luci cosiddette "seguipersona" all'interno di uno spettacolo: si dirigono, proiettando un fascio di luce, sul personaggio che è in scena e su cui deve essere posta l'attenzione, mentre il resto della scena seppur presente, rimane al buio. Allo stesso modo la coscienza sembra avere la funzione di "illuminare" e mantenere "accesa" quella parte della situazione che in quel momento è utile a quel determinato fine. La coscienza in questa prospettiva gioca un ruolo differente, ma non per questo meno importante, rispetto a quello giocato in un'architettura di tipo modulare. Senza di essa non sarebbe possibile far permanere un oggetto mentale. In sintesi possiamo dunque dire che la coscienza ha un ruolo fondamentale non tanto nell'integrare le informazioni (così come succedeva nel caso della modularità), quanto piuttosto nel permettere la *permanenza* e la *stabilità* dell'oggetto mentale. Senza di essa le immagini mentali rimarrebbero oscure, così come rimarrebbe oscura la scena teatrale, senza quelle luci che individuano ed illuminano il personaggio principale.

Capitolo 3

Approfondimento critico

3.1 Mental files e belief file: un confronto tra la posizione di Kovacs e quella di Perner

Una posizione alternativa a quella di Kovacs (2015) che abbiamo analizzato nel precedente capitolo, è quella di Perner (2015). L'idea di base è sempre la stessa: le abilità di *mindreading* sono possibili grazie alla capacità dell'individuo di reperire le informazioni provenienti dall'esterno e di inserirle all'interno dei cosiddetti file mentali.

La posizione di Perner e quella di Kovacs, che ad un primo sguardo possono sembrare simili, a ben guardare differiscono su molti punti. Innanzitutto sullo scopo che si prefiggono: Kovacs (2015) prova a trovare una soluzione in grado di spiegare in che modo, in maniera veloce e spesso automatica, riusciamo ad affrontare situazioni in cui sono coinvolte abilità di metarappresentazione. Perner, invece, si interroga sul modo in cui avviene nella nostra mente la rappresentazione delle credenze altrui e sulle possibili motivazioni che spiegherebbero la difficoltà nel riuscire in compiti di falsa credenza.

Se si analizzano le due posizioni, si arriva facilmente a dedurre, inoltre, che mentre Kovacs (2015) utilizza i *belief file* e gli *empty belief file*, per arrivare a sostenere che questi ultimi, da soli, rendono possibile la percezione precoce delle intenzioni altrui, Perner utilizza l'idea di *mental files* per sottolineare l'impossibilità di cogliere precocemente gli stati mentali altrui. Entrambi utilizzerebbero, dunque, lo stesso costrutto teorico per arrivare a sostenere tesi non solo differenti, ma opposte.

Per comprendere in che modo questo avvenga proviamo a confrontare le affermazioni dei due studiosi in proposito. Partiamo da una affermazione della Kovacs riguardo alla capacità di comprendere gli stati mentali altrui:

[...] the mere presence of social agents is sufficient to automatically trigger online belief computations not only in adults, but also in 7-month-old infants.[...] Hence, from 7 months on, an age by which infants attribute goals and intentionality, humans automatically compute other's beliefs and seem to hold them in mind as alternative representations of the environment. (Kovacs,2010: 1834).

Kovacs (2010) sta sottolineando due caratteristiche fondamentali del suo modello teorico: la precocità e l'automaticità. Sta infatti dicendo che a partire dai sette mesi i bambini sarebbero in grado di attribuire stati intenzionali in maniera automatica e che questa abilità è la stessa che permette agli adulti di computare e interpretare le credenze altrui. Secondo Kovacs a garantire questa precocità è proprio l'utilizzo dei *mental file*, in particolare degli *empty belief file*. Questi ultimi, infatti, permetterebbero di criticare l'idea, abbastanza diffusa in letteratura, che sia possibile rispondere positivamente a compiti (*impliciti*) di teoria della mente, utilizzando regole comportamentali di tipo associativo percettivo, e che dunque si possa parlare di computazione di credenze (*belief computation*) e di rappresentazione di credenze (*belief representation*) solo per riferirsi a compiti *espliciti* di teoria della mente:

However, according to some recent proposals, young infants tested with implicit ToM paradigms do not perform belief computations but instead form three-way associations between the agent, the object and the location

(Perner and Ruffman 2005), and recruit a so-called minimal theory of mind (Butterfill and Apperly 2013). It has been argued that while explicit ToM tasks involve a fully-fledged ToM that operates on belief representations, implicit tasks recruit a minimal ToM that relies on some simpler states that are not belief representations but make it possible to encode relations between the agent and its environment (Butterfill and Apperly 2013). If experimental evidence supports the possibility that infants are able to deal with empty belief files (where a relational encoding is not possible), such findings would suggest that core ToM components are present very early on in development, and that the basic representational structures used for belief tracking may be shared by infants and adults. (Kovacs, 2015: 11)

È chiaro dunque, che attraverso gli *empty belief file*, Kovacs (2015) vuole arrivare a sostenere che anche nel caso in cui non fosse possibile creare le associazioni (tra agente, oggetto e luogo) chiamate in causa da Perner, Ruffman (2005), sia possibile superare compiti in cui siano richieste abilità di *mindreading*. A tal proposito propone un esperimento che testimonierebbe in favore della sua teoria:

Consider, for instance, the earlier empty belief attribution situation, now adapted for infants: An infant is watching Agent 1 perform an invisible hiding event, putting a toy into one of two opaque boxes without the infant being able to see in which one. Then Agent 1 leaves the scene, and Agent 2 reveals the location of the toy, and she performs a second invisible hiding. Would young infants be able to infer the belief of Agent 1 when they themselves have no knowledge about the actual location of the object? Would they be able to compute at the moment of the first invisible hiding that the agent

believes the object to be somewhere (open an empty belief file) and fill in its content later, when it is revealed by a different agent? In such cases, three way associations are excluded (infants never see the agent, object, and location together). (Kovacs,2015: 11)

É possibile a questo punto comprendere perché l'utilizzo dei mental file da parte di Perner (2015) e da parte di Kovacs (2015) conduca a conclusioni differenti. Per Perner è fondamentale comprendere il contenuto della credenza per Kovacs no. In che modo secondo Kovacs (2015) è possibile computare e rappresentare credenze prescindendo dal loro contenuto? La risposta risiede nel diverso modo di intendere il processo che permette di superare compiti di falsa credenza. Kovacs (2015) individua due modalità attraverso le quali è possibile rappresentarsi e computare credenze: *post hoc inferences* e *continuous belief tracking*.

Quando ci troviamo di fronte ad una tipica situazione di falsa credenza è possibile che noi computiamo una credenza solo nel momento in cui ci viene chiesto (*on demand*), e per fare ciò dobbiamo richiamare alla mente tutta la sequenza di azioni che compiono i personaggi della scena (*post hoc inferences*). Se invece interpretiamo *continuamente (on line)* la sequenza di azioni che compiono i personaggi, non abbiamo bisogno di dover compiere questo dispendioso ritorno al passato per essere capaci di comprendere le credenze degli altri personaggi:

The differences between post hoc inferences and continuous belief tracking pertain to when, for what purpose, and possibly in how, the belief is computed. In the case of post hoc belief inference, one is cued by a question, or by the reappearance of the agent, to infer her beliefs based on earlier events. In these cases one must recall the details encoded about the scene, analyze past events (i.e., the temporal order of the event sequences, the role

different protagonists played in these events, whether they were present or absent during the most important scenes), and compute the agent's beliefs post hoc based on this information. In the continuous belief tracking scenario, in contrast, one spontaneously monitors what other people know and believe at all stages of the events. While continuous belief tracking refers to online and spontaneous computations of others' beliefs that happen under serious time pressure and it likely involves specific inferential processes, it is unclear whether these inferences are qualitatively different from post hoc belief inferences. (Kovacs,2015: 3)

Per riprendere la terminologia utilizzata da Kovacs, potremmo dire che Perner fonda il processo di attribuzione di credenze su *post hoc inference* per questo motivo non riesce ad attribuire ai bambini di età inferiore ai quattro anni la capacità di rappresentarsi credenze, al contrario di Kovacs (2015) che concepisce l'attribuzione di credenze come un processo continuo e costante. Ciò che fa Perner è insomma trovare una spiegazione a ciò che ritiene sia un'abilità complessa (quella di metarappresentazione), e lo fa utilizzando un costrutto teorico quello dei *mental file* sottolineando gli aspetti che riflettono tale complessità. Kovacs utilizza lo stesso costrutto per far emergere la precocità e dunque la semplicità delle medesime abilità.

Approfondiamo adesso l'utilizzo dei *mental file* da parte di Perner. I mental file, in particolare l'incapacità di collegare mental file tra loro, sono responsabili della tardiva comparsa di abilità di metarappresentazione del bambino. Secondo Perner, infatti, le abilità di teoria della mente sorgono in parallelo alla comprensione dell'aspetto intensionale della realtà che è percepito dal bambino tra i 4 e i 6 anni, dunque ad uno stadio più maturo del suo sviluppo. È solo in questa fascia di età che il bambino è già cosciente che può esistere una intenzionalità diversa dalla propria:

There is a curious window in child development, which opens when a child first passes verbal false belief tasks around 4 years and closes 2 years later when she passes second order belief tasks. During this period children appreciate that others can have beliefs that differ from their own. [...]Yet they seem to misunderstand the intensionality of belief, i.e., that belief about an object depends on the label under which the object is known to the believer. Understanding of intensionality appears with passing second-order belief tasks. (Perner,2015a: 77-78)

La posizione di Perner si fonda su un'argomentazione molto articolata che prende spunto, ed utilizza massicciamente l'idea di *mental files* così come formulata da Recanati (2012), da cui riprende anche l'articolata organizzazione strutturale dei file mentali adattandoli al problema della comprensione della falsa credenza. Vediamo innanzitutto in che modo Perner definisce i *mental files*:

A mental file is a tool for managing information about an object in the world (say, the file's referent or external referent). Files capture the predicative structure of language and thought: the distinction between what one is thinking/talking of (information that fixes the file's referent; individuating information) and what one thinks/says about it (the information about the referent on the file; predicative information) (Perner,2015a: 78-79)

I *mental file* sono strumenti che collegano la nostra mente ai dati della realtà sensibile e che ne permettono l'elaborazione e l'immagazzinamento. Se la spiegazione di Perner si fermasse qui, la sua posizione sarebbe molto simile a quella di Kovacs (2015). In realtà la funzione che Perner attribuisce ai *mental file* è molto più importante. Questi ultimi, che agiscono sulla struttura (predicativa) del linguaggio e del pensiero, permettono alla nostra mente di comprendere il rapporto tra l'entità concreta che deve essere percepita (il referente del file,

fissato da una data informazione identificativa), e tutte le sue possibili e ulteriori categorizzazioni (l'informazione predicativa sul referente contenuta nel file). Per arrivare a formulare la sua ipotesi (che la comprensione della coreferenzialità e la comprensione dell'intenzionalità altrui sono parti di una stessa abilità e si presentano insieme nel bambino), Perner parte da alcune considerazioni sull'idea di prospettiva:

In order to find a common denominator for understanding multiple identities and understanding false belief we propose that both rest on an understanding of perspective. (Perner,2002: 2)

L'adozione di una prospettiva coinvolge: aspetti visivi (vedere qualcosa da differenti angolazioni), aspetti mentali (per superare i compiti di falsa credenza devo essere in grado di adottare la prospettiva di un'altra persona), aspetti concettuali (una parte della psicolinguistica crede che categorizzare un oggetto in un modo o in un altro significa osservarlo da differenti prospettive).

Stando a quest'ultimo aspetto:

For, asked about a group of three rabbits and two cats, whether there are more rabbits or more animals, one has to count the rabbits as rabbits and as animals. Our analysis claims that two perspectives are involved. (Perner,2002: 19)

Un'altra situazione che Perner presenta, in cui appare molto chiaro in che modo i *mental file* lavorano in caso di coreferenzialità, è la seguente¹⁷: immaginiamo una storia in cui ci viene presentato un uomo come un vigile del fuoco. Verrà dunque aperto, secondo questa teoria, un mental file che rappresenta quest'uomo in quanto vigile del fuoco. Nel frattempo qualcuno trova una borsa e ci dice che questa borsa è di Mr Müller. Si crea dunque un nuovo file che contiene informazioni su Mr Müller. Verranno aperti dunque due file perché non siamo ancora a conoscenza del fatto che Mr Müller è il vigile del fuoco. Fino a quando non collegheremo i due file (con le rispettive informazioni) non sapremo mai che Mr Müller e il vigile del fuoco sono la stessa persona.

Ciò che Perner sta dicendo è che: a) automaticamente saremmo portati ad attribuire più facilmente a nomi/descrizioni differenti entità differenti e b) l'abilità di comprendere che due oggetti si trovano in un rapporto di coreferenzialità non si sviluppa così precocemente nel bambino, essendo una abilità più complessa.

Nella prima fase del processo di elaborazione dell'informazione che porterà all'adozione di una prospettiva (che sia visiva o mentale) si crea un legame tra oggetto e file mentale:

We say that a mental file is anchored to the object it collects information about. [...] A file is anchored to an external object, which individuates the object as a butterfly (not as an insect or an animal) and stores the predicative information that it is orange, etc. (Perner,2015a: 79)

¹⁷ L'esempio è tratto da un intervento di Perner a Torino durante l'EuroAsianPacific Joint Conference on Cognitive Science (EAP cogsci 2015).

La seconda fase del processo, è la più complessa ed è anche quella che permette la riuscita nei compiti di falsa credenza di secondo ordine¹⁸. Secondo Perner, proprio in virtù di quella definizione di prospettiva analizzata in precedenza, la comprensione degli aspetti che riguardano la sfera concettuale (cioè la comprensione che uno stesso oggetto può essere concettualizzato in maniera differente) coinvolge le stesse abilità messe in gioco nella comprensione degli aspetti che riguardano le abilità mentali. Così categorizzare un oggetto significa vederlo da una specifica prospettiva concettuale, e categorizzare lo stesso oggetto in maniera differente significa inserirlo in una differente prospettiva, quindi in un ulteriore file:

When a new label for an object is used the object is individuated in a new way, which creates a new conceptual perspective on that object. For each new conceptual perspective on an object, a new mental file is opened that records information that is available about the external referent from that perspective. [...] Each file individuates its referent in exactly one way. But it contains all predicative information that the files owner knows about the referent from that conceptual perspective. (Perner,2015: 79)

Per portare a termine la seconda fase del processo, i bambini devono essere in grado non solo di creare e aprire i mental file, ma anche di collegarli tra di loro. A tal proposito, Perner introduce, sempre prendendolo in prestito da Recanati, l'idea di collegamento (*linking*) tra file mentali:

Linking is an operation of connecting different files so that the information contained in each file can flow freely to any other file linked to it (Perner,2015b: 11)

¹⁸ È importante notare che per Perner è soltanto dopo che il bambino riesce a superare questi test che si può parlare di comprensione dell'intenzionalità altrui.

L'idea che i file siano collegati tra loro è condizione necessaria affinché l'informazione sia resa disponibile. Allo stesso tempo l'idea che un file possa scorrere liberamente verso gli altri file ad esso collegati rischia di rendere il sistema confuso e poco efficace. È necessario dunque stabilire che tipo di file entrano in collegamento e con che criterio un file si collega all'altro.

Chiariamo innanzitutto che il tipo di collegamento più primitivo che è necessario operare, è quello tra i *vicarious file* e il *regular file*. I primi hanno bisogno di essere collegati ai secondi per fare in modo che la rete di file funzioni:

Vicarious files are linked to regular files in order to establish sameness of referent between linked files. However, the flow of information between linked regular and vicarious files must be constrained, since a vicarious file may contain misinformation (e.g., the chocolate still being in its old location), which must be quarantined and not contaminate the information in the regular files. (Perner,2015a: 79)

Potremmo definire i *vicarious file* come i corrispondenti *empty belief file* teorizzati da Kovacs (2015). Sono infatti degli strumenti che permettono, se collegati ai *regular file*, di concepire l'esistenza di una prospettiva diversa dalla propria. Potrebbero essere definiti insomma come file che non contengono credenze (pienamente) accolte dal soggetto come proprie. Nel caso di Kovacs, però, grazie alla sola apertura degli *empty belief file*, era possibile riuscire in compiti di metarappresentazione, nel caso di Perner, ed è qui la grande differenza tra i due

teorici dei mental file, l'apertura di un *vicarious file* non implica di per sé alcun tipo di abilità particolare. Vediamo in che modo è possibile descrivere il rapporto tra *vicarious file* e *regular file*. Per fare ciò immaginiamo di trovarci di fronte alla classica situazione presentata nei test di falsa credenza in cui abbiamo un soggetto che assiste alla celebre scena:

[...] Suppose Susi sees a ball being put in box 1; in her absence the ball is moved to box 2. How does a test subject who observes this situation store information about Susi's beliefs? For mature observers a regular file for the ball shows it in its new location while a vicarious file indexed to Susi shows it in its old location. (Perner,2015a: 79)

Il *vicarious file* è rappresentato dalla credenza che Susi crea nella sua mente: l'oggetto si trova nel luogo in cui lei lo ha visto l'ultima volta. Ma lo spettatore che ha visto la scena sa bene che l'oggetto è stato spostato dal posto in cui l'ha visto Susi, crea dunque un *regular file*. Dopo aver creato il file lo spettatore deve, secondo Perner, rendersi conto che sia lui che Susi si stanno riferendo allo stesso oggetto e che dunque i due file hanno lo stesso referente esterno. Quando il *vicarious file* sarà collegato al *regular file* lo spettatore supererà il test di falsa credenza e si potrà dunque affermare che è capace di comprendere che esiste una intenzionalità diversa dalla propria.

In sostanza l'apertura del *vicarious file* segnala l'esistenza di una differente prospettiva, che sarà percepita solo dopo che i file saranno collegati. Anche in questo caso Perner (2015) utilizza una nozione utilizzata da Recanati(2012):

[...]in belief ascriptions, the files associated with linguistic occurrences do not necessarily reflect the speaker's current point of view, but may reflect the ascriber's point of view. In other words, attitude ascriptions allow files to be used vicariously. This introduces us to the metarepresentational function of mental files. Mental files are primarily singular terms in the language of thought: they serve to think about objects in the world. But they have a derived, metarepresentational function: they serve to represent how other subjects think about objects in the world. (Recanati,2012: 147)

All'incapacità di collegare file che condividono lo stesso referente segue l'incapacità di comprendere l'intenzionalità altrui. La capacità di comprendere che a differenti concettualizzazioni può corrispondere uno stesso referente esterno, permette di assumere una nuova prospettiva. Questa abilità è necessaria per riuscire nei compiti di falsa credenza di secondo ordine:

In sum we can state that for a proper understanding of belief one needs the ability to link files that share the same referent. (Perner,2015b: 13)

Ovviamente il *vicarious file*, può contenere una credenza non vera e dunque compromettere i successivi legami tra file, per questo motivo è necessario isolare quel file:

However, the flow of information between linked regular and vicarious files must be constrained, since a vicarious file may contain misinformation (e.g.,

the chocolate still being in its old location), which must be quarantined and not contaminate the information in the regular files. For this reason vicarious files are linked to regular files by vertical links (Recanati, 2012, chap. 14 and 15), which do not enable the transfer of information as freely as horizontal links. (Perner,2015a: 79)

É necessario dunque che il collegamento tra file non sia del tutto libero, ma segua dei criteri, che riprendendo la terminologia di Recanati, Perner chiama “*vertical linking*” e “*horizontal linking*”. Il legame tra *vicarious* e *regular file* deve essere verticale in quanto il trasferimento delle informazioni non può essere del tutto libero:

[...] Vicarious files are linked to regular files by vertical links, which do not enable the transfer of information as freely as horizontal links (Perner,2015: 79)

Il legame orizzontale è invece riservato agli elementi che possono essere collegati liberamente:

[...] horizontal links between files that are anchored to the same referent enable free flow of information between linked files. Horizontal links establish sameness of referent between linked files. [...] Each file captures a different conceptual perspective on the object. (Perner,2015a: 79)

3.2 *Dual systems e multi-systems: due modelli a confronto*

Michael & Christensen (2016) hanno recentemente pubblicato un articolo in cui criticano ancora una volta l'approccio di Apperly & Butterfill.

L'idea che siano due sistemi (differenti qualitativamente) a rendere possibile il processo di attribuzione di credenze, continua a non convincere i teorici della cooperazione tra sistemi. In particolare, a non convincerli è che sia possibile suddividere il processo che porta alla rappresentazione di stati mentali in due parti differenti. Nella prima parte del processo (dunque grazie al primo sistema) sarebbe possibile attribuire scopi. Nella seconda parte sarebbe possibile comprendere stati psicologici. In sostanza non trovano plausibile che l'attribuzione di scopi possa avvenire in una concezione così minimalista come quella ipotizzata nel primo sistema. Secondo Michael & Christensen se è vero che il modello di Apperly & Butterfill avrebbe il vantaggio di fornire una spiegazione ai risultati discordanti presenti in letteratura tra competenze di teoria della mente implicite ed esplicite, è anche vero che il modello così concepito finirebbe per risultare implausibile.

Michael & Christensen (2016), in particolare, si soffermano su due aspetti che a loro avviso rendono il modello duale poco convincente: il modo in cui viene descritto il processo di attribuzione di scopi (che deriva dalle caratteristiche che attribuiscono a questi ultimi) e l'architettura cognitiva che ne pongono alla base, modulare e dunque incapsulata. Vediamo perché:

Butterfill and Apperly's characterization of goal attribution is designed to show how goals could be ascribed by infants without representing them as related to other psychological states, and the minimal mindreading system is supposed to operate without employing flexible semantic executive cognitive processes. But research on infant goal attribution reveals that infants exhibit a high degree of situational awareness that is strongly suggestive of flexible semantic executive cognitive processing, and infants appear moreover to be sensitive to interrelations between goals, preferences and beliefs. (Michael & Christensen, 2016: 2)

Ciò che stanno dicendo Michael & Christensen è che Apperly & Butterfill non forniscono prove sufficienti che testimonierebbero in favore della possibilità di poter prescindere dalla comprensione degli stati mentali nel processo di attribuzione degli scopi. Per abbracciare la concezione minimalista Apperly & Butterfill sono costretti infatti a ipotizzare l'esistenza di un processo intermedio per arrivare alla comprensione di stati mentali.

La possibilità di Apperly & Butterfill di ipotizzare questo stato intermedio deriva ancora una volta dalla concezione di scopo (poco esplicativa, secondo Michael & Christensen), a cui fanno riferimento. Quest'ultimo è semplicemente definito come il risultato dell'azione alla quale è diretto ("goal is an outcome to which an action is, or will be, directed"). Questa definizione non convince Michael & Christensen, infatti:

goals are represented as outcomes which are functionally-teleologically related to actions – it doesn't describe the specific form of representation employed to attribute or reason about goals." (Michael & Christensen, 2016: 9)

L'idea che lo scopo sia legato "funzionalmente all'azione" non è sufficiente a dare credito alla visione minimalista che Apperly & Butterfill intendono sostenere. Soprattutto non risulta chiaro cosa contribuisca, se non la percezione degli stati psicologici, alla attribuzione e alla possibilità di pensare agli scopi stessi:

Nor is any clear account given of the conditions in which goals are attributed and the kinds of information that contribute to goal attribution, other than the proscription on information concerning psychological states. B&A only say that there is evidence that young children, non-human primates, and corvids can track the functions of things (2013, p. 614). But a positive theory of goal attribution must specify how infants identify particular outcomes as the goals of actions. (Michael & Christensen, 2016: 9)

Il tentativo di voler separare l'attribuzione di scopi dagli stati psicologici, non sembrerebbe dunque essere supportata da sufficienti prove empiriche. La letteratura in proposito sembrerebbe infatti suggerire, al contrario, una interrelazione tra questi due stessi aspetti dello stesso processo:

A primary motivation for B&A's teleological account of goal attribution is to avoid the need to make the representation of goals dependent on other psychological states, like desires and beliefs, yet this evidence suggests that infants do treat goals, preferences and beliefs as interrelated and mutually influencing. (Michael & Christensen, 2016: 19)

Per quanto riguarda invece il secondo punto, e cioè l'incapsulamento informativo, non è difficile ipotizzare che secondo Michael & Christensen

(2016), sostenitori della cooperazione tra sistemi, non sia ipotizzabile che l'attribuzione di scopi e credenze avvenga in una situazione di impermeabilità tra conoscenze differenti. Ciò che propongono ad Apperly & Butterfill è infatti l'abbandono della struttura modulare e incapsulata in favore dell'adozione di una visione in cui più sistemi specializzati interagiscono tra loro attraverso l'utilizzo delle funzioni esecutive:

The second type of response that B&A might adopt is to abandon the idea that infant mindreading is strongly encapsulated. Instead, mindreading might involve an interplay in which a variety of specialized systems are integrated via executive cognition. Thus, specialized systems for causal representation, agent tracking, the representation of agent's attitudes, and action structure might all take input from and contribute to generalized situation awareness. (Michael & Christensen,2016:24)

Butterfill & Apperly (2016) in un articolo di risposta alle critiche si occupano di difendere la loro posizione non accettando le due possibili soluzioni che Michael & Christensen (2016) avevano offerto alla fine della loro argomentazione:

M&C offer two potential ways in which we might respond to their critique: either we must abandon the idea that minimal mindreading is strongly encapsulated and accept that it might consist of multiple systems which require integration via the use of executive functions, or else we must appeal to 'ad hoc representations' in offering an account of cognitively efficient goal ascription. (Butterfill & Apperly,2016: 3)

Butterfill & Apperly (2016) spiegano in che modo il processo di attribuzione di scopi è possibile anche in una concezione minimalista, ed inoltre chiariscono i termini entro i quali l'incapsulamento che regola i due sistemi è, secondo loro, non solo possibile ma necessario.

Quando ipotizzano che il *minimal mindreading* sia incapsulato non intendono sostenere che il funzionamento di quest'ultimo, non sia influenzato da altri tipi di conoscenze o dallo sviluppo delle funzioni esecutive. Per chiarire in che modo questo sia possibile prendono ad esempio il caso della cognizione numerica nel bambino. L'abilità di enumerare oggetti agisce in concomitanza con altre abilità, non ultima quella di elaborare strategie per cercare di selezionare il più ampio numero di oggetti possibile. Ma ciò che fanno notare Butterfill & Apperly (2016) è che per quanto queste capacità (che collateralmente appaiono nel bambino) possano svilupparsi, i bambini non riescono mai ad enumerare più di tre oggetti:

For our current purposes this case is helpful as an illustration of how a relatively encapsulated process comes to enable a growing range of responses and an increasingly rich set of judgements over development. At least some of these developments are likely to depend on infants' increasing knowledge and increasing executive function, of course. But note that this does not affect the ability to enumerate sets of up to 3 objects, whose signature limit of 3-4 items remains unaffected by these other developments. (Butterfill & Apperly, 2016: 5)

Una situazione simile a quella della cognizione numerica potrebbe riguardare dunque anche il *minimal mindreading* che potrebbe servirsi per

il suo funzionamento di altre risorse cognitive come la working memory e il controllo inibitorio:

A similar situation may exist for minimal mindreading abilities. These abilities are cognitively efficient in the sense that, compared with full-blown mindreading, they trade some flexibility for reduced demands on scarce cognitive resources such as working memory and inhibitory control. (Butterfill & Apperly, 2016: 6)

Per quanto riguarda invece la critica riguardo al modo non convincente con cui Apperly & Butterfill descrivono il processo di attribuzione degli scopi, la questione non è così semplice come potrebbe apparire dalla descrizione che Michael & Christensen (2016) fanno della concezione minimalista. Partiamo dalla definizione di *goal ascription* che forniscono Butterfill & Apperly (2016):

Goal ascription is the process of identifying an outcome to which an observed or anticipated sequence of bodily configurations and joint displacements are directed. How could goal ascription sufficient to explain success on a range of false belief tasks be cognitively efficient? (Butterfill & Apperly, 2016: 10)

Questa descrizione del processo di *goal ascription* (strettamente legata alla percezione delle configurazioni corporee) è supportata da una letteratura che considera processi e rappresentazioni motori coinvolti nell'attribuzione di scopi:

One strand of existing research on this topic hinges on the idea that motor processes and representations are involved in some cases of goal ascription. Control of action involves motor representations of outcomes that are relatively distal from bodily configurations and joint displacements, outcomes such as the grasping of a handle or the movement of an object from one place to another. (Butterfill & Apperly,2016: 10)

Inoltre le rappresentazioni motorie di uno scopo possono avvenire non solo quando un soggetto compie un'azione, ma anche quando la osserva:

Importantly, motor representations of an outcome can occur not only when agents are performing an action directed to that outcome but also when they are passively observing such an action (Rizzolatti and Sinigaglia 2010). (Butterfill & Apperly,2016: 10)

Affidarsi alle rappresentazioni motorie per giustificare il processo di attribuzione degli scopi è un modo che permette a Butterfill & Apperly (2016) di rispondere in maniera pertinente alle critiche mosse dai teorici della cooperazione tra sistemi infatti:

The conjecture that motor representations and processes underpin a kind of goal ascription is relevant to us because such processes enable rapid, online action predictions, indicating that they may be cognitively efficient in the sense required for minimal mindreading. (Butterfill & Apperly,2016: 10)

In questo modo Butterfill & Apperly (2016) giustificano la presunta non causalità attribuita alla loro concezione minimalista.

3.3 Bayesian change-analysis: i limiti del metodo statistico tradizionale

In un recente articolo Baker, Leslie, Gallistel, Hood (2016), espongono un nuovo punto di vista riguardo alle metodologie utilizzate per valutare la presenza di abilità di teoria della mente. Propongono infatti l'utilizzo di un metodo statistico di tipo bayesiano, il "bayesian change-point analysis":

The Bayesian change-point analysis employed here differs from traditional techniques in developmental research in two main ways. First, it focuses on change at the level of the individual. [...] Because learning, like development, takes place in the individual brain, understanding of the acquisition process should be based on individual, not group, curves. The second main difference between the Bayesian changepoint analysis and traditional techniques in developmental research is that, as the name suggests, it uses Bayesian statistics, rather than tNHST¹⁹. In contrast to tNHST, Bayesian change point analysis provides a tool for evaluating the strength of the evidence in favor of various models of change in the probability of obtaining a value in a chronological record (Dienes, 2011; Lee & Wagenmakers, 2005; Salsburg, 2001). (Baker et.al.2016: 129)

¹⁹ Si riferisce al tipo di test più frequentemente utilizzati in psicologia dello sviluppo.

La differenza generale tra i modelli statistici tradizionalmente utilizzati in psicologia dello sviluppo e il metodo bayesiano è che quest'ultimo concepisce il risultato non come qualcosa di definito e unitario, quanto piuttosto come un processo il cui risultato finale è proprio il processo stesso. In questo senso se applicato al caso della misurazione delle abilità di teoria della mente, ciò a cui mira il metodo in questione non è tanto comprendere a che età queste abilità abbiano origine, quanto piuttosto quale andamento segue il loro sviluppo. Questa analisi avviene avendo come riferimento il cambiamento all'interno dell'individuo.

Ciò che è emerso lavorando ai dati raccolti utilizzando questo metodo è che non è del tutto corretto descrivere i risultati sui test di falsa credenza caratterizzando l'esito con la dicotomica divisione superamento/fallimento:

Performance at level of an individual child is not binary and we cannot always categorize a child in terms of "failing" or "passing". (Baker et.al.,2016: 140)

Il dibattito sull'età in cui tali abilità si manifestano, che è stato analizzato nei precedenti paragrafi, assume dunque una nuova luce. Ciò che vengono calcolate sono le probabilità che tali abilità si verifichino nell'intervallo di tempo considerato, anche perché stabilire cosa significa superare o fallire il test attraverso questa metodologia diventa una questione difficile. Vediamo perché:

However, it does seem to us that one upshot of these data is that the traditional notion of passing and failing the false belief task (and perhaps other developmental tasks) has limited value. Rather we suggest it is better to think in terms of the probability of passing: what changes with development is the probability of passing the false belief task, and it changes in different ways in different children at different times. Perhaps this fits better with a complex of performance systems whose interactions fluctuate daily but which become gradually better integrated, rather than with a single change in an underlying competence. From a practical point of view, a single trial recorded as a binary response does not reveal much about underlying states. (Baker et.al.,2016: 143)

Ciò che cambia con lo sviluppo del bambino, in questa prospettiva, è la probabilità di superare i compiti di falsa credenza.

In accordo con l'idea, ormai largamente condivisa in letteratura, che il superamento dei test di falsa credenza coinvolga differenti abilità, e che dunque parlare di abilità di teoria della mente significhi analizzare una serie di abilità ad essa sottostanti, questo metodo sottolinea la difficoltà di porre la questione nei termini di superamento/fallimento dei test.

É importante notare come l'utilizzo di questa metodologia non abbia lo scopo di sostituire i metodi statistici tradizionali, quanto piuttosto di arricchirli. In questo senso il bayesian-change point analysis colma un vuoto presente in letteratura, che riguarda le modalità attraverso le quali le abilità di teoria delle mente si presentano nell'individuo e che tipo di sviluppo seguono:

For developmentalists, the difference between an individual passing and failing is of paramount importance. Indeed, one aim of developmental research is to describe and explain how change occurs. What is the form of the transition between the failing and the passing states? Is it “sudden insight” or “gradual change” or something else? Do all children present the same profile of change from failing to passing? How can we characterize variability within and between individuals? These are the types of developmental questions that Bayesian change point analysis can answer. (Baker et.al.,2016:126)

Conclusione

Attraverso l'analisi delle posizioni presenti in letteratura sulle abilità di teoria della mente e sulla flessibilità cognitiva è stato possibile cogliere in che termini le due abilità risultano estremamente legate. Ripercorrendo la letteratura sull'*early mindreading*, ci si è resi conto che la questione della flessibilità è centrale sia negli approcci di matrice innatista che in quelli *behaviour-rules based*. I primi, quando sostengono che i bambini in età precoce sono capaci di abilità di *mindreading*, presuppongono l'esistenza di una forma di flessibilità precoce. I secondi, gli approcci *behaviour-rules based*, invece, pur sostenendo di poter fare a meno della flessibilità cognitiva nella spiegazione delle abilità inferenziali dei bambini, in realtà la presuppongono. Sulla base di quanto detto finora si è rivelato necessario provare a fornire un'alternativa ai modelli presenti in letteratura. Provando a concepire alla base della percezione degli stati mentali un'architettura, nel nostro caso bayesiana, differente da quella modularista (paradigma per diverso tempo dominante in ambito psicologico), ne è venuto fuori un modello cognitivo in cui l'impermeabilità tra differenti tipi di conoscenza e la dinamicità con cui avviene lo scambio tra informazioni organizzate gerarchicamente, non lascia spazio alla ormai dispendiosa e poco esplicativa versione innatista/modularista né tantomeno ad una visione fondata su regole comportamentali *tout-court*. Anche il ruolo della coscienza in questa visione viene ridimensionato. Mentre nel modularismo era necessaria a integrare

informazioni provenienti da domini differenti, in una visione fondata sulle reti bayesiane quest'ultima assume la funzione di permettere la permanenza dell'oggetto mentale nel focus attentivo. Inoltre l'utilizzo della metodologia bayesiana applicata all'interpretazione dei test di falsa credenza oltre a far comprendere i limiti di un tipo di ricerca che mira ad individuare uno specifico momento dello sviluppo del bambino nel quale tali abilità si manifestano, ha messo in luce quanto le abilità di *mindreading* non siano del tipo tutto-o-nulla, e questo sembra essere un altro tassello in favore di una visione che vede il *mindreading* come una capacità che si costruisce nel tempo sulla base della generalizzazione dei dati provenienti dall'esperienza.

Bibliografia

Apperly, I. A., Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states?, *Psychological Review*, 116, 953-970.

Apperly, I.A., Butterfill, S. A. (2013). How to construct a minimal theory of mind, *Mind & Language*, 28(5), 606-637

Austerweil, J.L., Gershman, S.J., Tenenbaum, J., Griffiths, T. (2015). *Structure and Flexibility in Bayesian Models of Cognition*, in Busemeyer, J., Townsend, J., Wang, Z., Eidels., Oxford Handbook of Computational and Mathematical Psychology, Oxford University Press

Bandhari, A., Duncan J. (2014). Goal neglect and knowledge chunking in the construction of novel behavior, *Cognition*, 130 (1), 11-30

Botvinick, M.M., Niv, Y., Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, 113, 262–280

Butterfill, S.A., Apperly, I.A. (2013). How to construct a minimal theory of mind, *Mind & Language*, (28), 5, 606-637.

Butterfill, S.A., Apperly, I.A. (2016). Is Goal Ascription Possible in Minimal Mindreading?, *Psychological Review*, 123 (2), 228-233.

Carruthers, P. (2006). *The architecture of the mind. Massive modularity and the flexibility of thought*, Clarendon press, New York, Oxford University Press

Carruthers, P. (2013). Mindreading in infancy, *Mind & Language*, 28 (2), 141-172

Christiansen, W., Michael, J. (in press) From two systems to a multi-systems architecture for mindreading, *New Ideas In Psychology*

De Bruin, L.C., Newen, A. (2012) An association account of false belief understanding, *Cognition*, 123(2), 240–259

Dehaene, S.J.R., King (2016) Decoding the Dynamics of Conscious Perception: The Temporal Generalization Method, *Micro-, Meso- and Macro-Dynamics of the Brain*

Diaconescu, A., Mathys, C., Weber, L., Daunizeau, J., Kasper, L., Lomakina, E., Fehr, E., Stephan, K. (2014) Inferring on the Intentions of Others by Hierarchical Bayesian Learning, *Computational Biology*, 10, 1-19

- Diamond,A.(2013). Executive Functions. *Annual Review of Psychology*, 64, 135-168.
- Duncan, J. et.al. (2000). A neural basis for general intelligence, *Science*,289, 457-460
- Duncan,J.et.al.(2008). Goal Neglect and Spearman's g: Competing Parts of a Complex Task, *Journal of Experimental Psychology*,137,1, 131–148
- Duncan, J. (2010). *How intelligence happens*, Yale University Press
- Duncan,J.(2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behavior, *Trends in cognitive sciences*, 14,172-179
- Duncan,J.(2013). The Structure of Cognition: Attentional Episodes in Mind and Brain, *neuron*(80), 35-50
- Eco, U., (1984). *Semiotica e filosofia del linguaggio*, Einaudi, Torino
- Gelhen (1940). *Der Mensch. Seine Natur und seine Stellung in der Welt*, Junker und Dünnhaupt, Berlin [trad.it., *L' uomo. La sua natura e il suo posto nel mondo*, Feltrinelli, Milano 1983]
- Goldberg, E (2001). *The executive brain. Frontal lobes and the civilized mind*, New York, Oxford University Press
- Gopnik, A. (1993). How we read our own minds: The illusion of first-person knowledge of intentionality, *Behavioral and Brain Sciences*, 16, 1–14.
- Gozzano, S. (2001). *Mente senza linguaggio*, Editori riuniti, Roma.
- Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., Tenenbaum, J. B. (2010) Probabilistic models of cognition: Exploring representations and inductive biases, *Trends in Cognitive Sciences*, 14, 357-364.
- Hassin, R.R., Bargh, J.A., Zimerman, S. (2009) Automatic and flexible: the case of non-conscious goal pursuit, *Social cognition*,27(1), 20-36
- Hassin, R.R., Aarts H., Eitam, B., Custers R., Kleiman,T. (2009). *Non-conscious goal pursuit, working memory, and the effortful control of behavior* in E. Morsella, J. A. Bargh & P. M. Gollwitzer (Eds.), *Oxford Handbook of the Psychology of Action* (pp. 549-568). New York, NY: Oxford University Press

Heyes, C. (2013). Submentalizing: I'm not really reading your mind, *Perspectives on Psychological Science*, 131-143.

Heyes, C. (2014). False belief in infancy: a fresh look. *Developmental Science*, 17, 647-659.

Herder, J.G. (1770). *Abhandlung über den Ursprung der Sprache*, Voss, Berlin [trad. it., *Saggio sull'origine del linguaggio*, Pratiche, Parma 1995]

Hutto, D., Herschbach, M., Southgate, V. (2011). Social cognition: Mindreading and Alternatives, *Review of Philosophy and Psychology* 2, (3), 375-395.

Karmiloff-Smith, A. (1992) *Beyond Modularity. A Developmental Perspective on Cognitive Science*, Cambridge, MIT Press

Kissine, M. (2012). Pragmatics, Cognitive Flexibility and Autism Spectrum Disorders, *Mind & Language*, 27, 1-28.

Kovacs, A., (2015). Belief Files in Theory of Mind Reasoning, *Review of Philosophy and Psychology*, 1-19

Leslie, A. M, Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism, *Cognition*, (43), 225-51.

Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies, *Trends in Cognitive Sciences*, 9(10), 459-462.

Martin, A., Santos, L.R. (2016). What cognitive mechanisms support primate theory of mind? *Trends in Cognitive Sciences*

Machery, E. (2008). Massive modularity and the flexibility of human cognition. *Mind & Language*, 23(3), 263-272

Martin, A., Santos, L. R. (2014). The origins of belief representation: Monkeys fail to automatically represent others' beliefs, *Cognition*, 130 (3), 300-308.

Mazzone, M. (2014). Mental states as generalizations from experience: a neurocomputational hypothesis. *Philosophical Explorations*, 17, 223-240

Mazzone, M. (2011). Schemata and associative processes in pragmatics, *Journal of pragmatics*, 43, 2148-2159

- Michael, J., Christensen, W. (2016) Flexible goal attribution in early mindreading. *Psychological Review*, 123(2), 219-227.
- Onishi, K.H., Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308, 255-258.
- Perner, J., Stummer, S., Sprung, M., Doherty, M. (2002). Theory of mind finds its Piagetian perspective: why alternative naming comes with understanding belief, *Cognitive Development* 103,1,22-19
- Perner, J. (2010). Who took the cog out of cognitive science? Mentalism in an era of anti-cognitivism. In P. Frensch, & R. Schwarzer (Eds.), *Cognition and neuropsychology: International perspectives on psychological science* (Vol. 1, p. 241-61), New York Psychology Press.
- Perner, J., J. Roessler (2012) From infants' to children's appreciation of belief, *Trends in Cognitive Sciences* 16(10), 519–525.
- Perner, J. (2014). Commentary on Ted Ruffman's "Belief or not belief:...". *Developmental Review*, 34, 294-299.
- Perner, J., Huemer, M., Leahy, B. (2015) Mental Files and Belief: A Cognitive Theory of How Children Represent Belief and its Intensionality, *Cognition*, 145, 77-88.
- Perner, J., Leahy, B. (2015) Mental Files in Development: Dual Naming, False Belief, Identity and Intensionality, *Review of Philosophy and Psychology*, 1-18.
- Povinelli, J., Vonk, J. (2003) Chimpanzee minds: suspiciously human? *Trend in cognitive science*, 7, 157-160
- Povinelli, J., Vonk, J. (2004) We don't need a microscope to explore the Chimpanzee's mind. *Mind & Language*, 9(19), 1-28
- Premack, D., Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4), 515-26
- Rakison, H., Lopyan, G. (2008) Developing object concepts in infancy: An associative learning perspective, *Monographs of the Society for Research in Child Development*, 73(1)

- Repacholi, B. M., Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds, *Developmental Psychology*, 33(1), 12–2
- Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental review*, 34, 265-293.
- Ruffman,T. Perner, J. (2005) Infants'insight into the mind: how deep?. *Science*,214-216.
- Ruffman,T. Perner, J. (2005) Do infants really understand false belief? *Trends in Cognitive Sciences*, 9(10), 462–463.
- Samson,D.(2013) *Theory of mind*. In D.Reisberg (Ed.). Oxford Handbook of cognitive Psychology ,943-956, Oxford University Press
- Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and 'theory of mind'. *Mind & Language*, 14(1), 131-153.
- Shanks,D.R.,(2010). Learning: from association to cognition, *Annual Review of Psychology*,61,273-301
- Shanahan, M.,Baars, B. (2005) Applying global workspace theory to the frame problem, *Cognition*, 98, 157–176
- Smith, L.B. (2000) Avoiding association when its behaviorism you really hate. In R. Golinkoff & K Hirsh-Pasek. (Eds.) *Breaking the word learning barrier*. Oxford:Oxford University Press, 169-174.
- Sperber, D. (2005) *Modularity and relevance: How can a massively modular mind be flexible and context-sensitive?* in *The Innate Mind: Structure and Content* P. Carruthers, S. Laurence, S. Stich
- Surian, L., Caldi, S., Sperber, D. (2007). Attribution of beliefs by 13- month-old infants, *Psychological Science*, 18(7), 580–586
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011) How to grow a mind: Statistics, structure, and abstraction. *Science*, 331, 1279-1285.
- Tomasello, M (2014) *Unicamente umano. Storia naturale del pensiero*, il Mulino, Bologna.
- Tomasello, M. Call, J. (2003) Chimpanzees versus humans: it's not that simple. *Trends in cognitive science*,7,157-160.

Tomasello, M. Call, J. (2008) Does the chimpanzee have a theory of mind? 30 years later *Trends in cognitive science*,12,187-19

Virno, P. (2002) *Scienze sociali e "natura umana". Facoltà di linguaggio, invariante biologico e rapporti di produzione*, Rubbettino 2002, Soveria Mannelli.