

APhEx 21, 2020 (ed. Vera Tripodi)
Ricevuto il: 14/01/2020
Accettato il: 20/01/2020
Redattori: Claudio Calosi & Pierluigi Graziani

APhEx
PORTALE ITALIANO DI FILOSOFIA ANALITICA
GIORNALE DI **FILOSOFIA**
NETWORK
N°21, 2020

L e t t u r e c r i t i c h e

Paul Dumouchel e Luisa Damiano, **Vivere con i robot. Saggio sull'empatia artificiale**, Raffaello Cortina, Milano, 2019, pp. 220.

Francesco Bianchini ed Edoardo Datteri

1. Introduzione

Vivere con i robot è qualcosa che sta avvenendo da un po' di tempo, che accadrà nel prossimo futuro, che diventerà reale nel lungo periodo. Queste tre affermazioni sono tutte vere a seconda della prospettiva che si assume rispetto all'avanzamento tecnologico della robotica e alla sua diffusione nella vita di tutti i giorni degli esseri umani. In particolare, tali affermazioni sono vere da un punto di vista sociale, nel senso che si assiste a una progressiva integrazione delle pratiche umane con gli artefatti robotici, integrazione che sempre più si sposta da un ambito produttivo e industriale a

uno interattivo tipico dei contesti sociali. Tale è il contesto all'interno del quale si situa il filone di ricerca presentato da Dumouchel e Damiano in *Vivere con i robot. Saggio sull'empatia artificiale*, testo ora tradotto in italiano ed edito da Raffaello Cortina, ma già pubblicato in francese dalle Editions de Seuil e in inglese da Harvard University Press.

L'interesse suscitato da queste tematiche è evidente e al tempo stesso pervasivo della nostra cultura, sempre più abituata a confrontarsi con i diversi mondi che ricadono all'interno dell'universo dell'intelligenza artificiale. Gli autori del libro, entrambi filosofi, declinano questo interesse in una prospettiva epistemologica, che non trascura i dibattiti che hanno animato, da un punto di vista teorico, gli ultimi decenni della scienza cognitiva e i molteplici approcci alla cognizione che si sono determinati nel tempo. Il tema centrale del libro resta, tuttavia, la robotica, e, nello specifico, un tipo particolare di robotica, la robotica affettiva che si inquadra nel più vasto campo, consolidatosi nell'ultimo decennio, della robotica sociale.

Un robot sociale è un artefatto robotico autonomo che interagisce con gli esseri umani o altri agenti del mondo fisico seguendo un comportamento più o meno regolato da direttive e da tratti che definiscono il suo ruolo. Per tale ragione, gli aspetti corporei, come ad esempio la presenza di caratteristiche senso-motorie, fanno sì che negli studi e nelle applicazioni robotiche assuma notevole rilevanza l'*embodiment* fisico dell'artefatto. La robotica sociale si inserisce perfettamente nel quadro teorico dell'*embodiment*, seppure non siano secondarie nella progettazione di un robot sociale le conoscenze relative alle abilità cognitive tipiche degli studi sulla cognizione sociale (Breazeal *et al.*, 2016). Inoltre, l'utilizzo di metodi simulativi richiede un'armonizzazione con la nozione di *embodiment* che viene chiamata in causa dal punto di vista sperimentale, come vedremo nel prossimo paragrafo.

In *Vivere con i robot* è trattata principalmente solo una particolare tematica relativa alla robotica sociale, quella che riguarda la robotica affettiva. A partire da una riflessione su alcune delle nozioni che definiscono l'intelligenza artificiale e la robotica, quali quella di autonomia, gli autori prendono in oggetto in larga parte la robotica umanoide, affrontando il tema dei robot come sostituti degli esseri umani in determinati compiti, l'autorità che essi sono chiamati a esercitare, l'ineliminabile richiesta della loro presenza e le questioni poste dalla zona perturbante, la cosiddetta ipotesi della *uncanny valley*, avanzata da Masahiro Mori ormai diversi decenni fa (Mori, 1970). A tali riflessioni gli autori fanno seguire la

tesi principale del libro, che consiste nell'idea che sia possibile parlare autenticamente di emozioni in ambito robotico superando simultaneamente la visione internista e quella esternista delle emozioni in ambito cognitivo, in vista di una prospettiva integrata imperniata su una visione sistemica e coordinata delle emozioni stesse, sia fra esseri umani che fra esseri umani e robot. A supporto di tale tesi Dumouchel e Damiano portano alcune riflessioni critiche sui vari approcci allo studio della mente che hanno caratterizzato la scienza cognitiva, l'intelligenza artificiale e la filosofia della mente negli ultimi cinquanta anni e una trattazione del metodo sintetico come filosofia sperimentale in robotica cognitiva da un punto di vista epistemologico. L'approccio sintetico viene trattato anche in una dimensione etica nella parte finale del volume, dopo un capitolo dedicato alle principali e più recenti applicazioni nell'ambito della robotica affettiva.

Al fine di valutare criticamente a che condizione per gli autori è possibile parlare di emozioni ed empatia artificiali, nei prossimi paragrafi discuteremo alcuni aspetti relativi al ruolo che hanno i robot sociali nella comprensione degli aspetti cognitivi e degli esseri umani in generale. Prenderemo in considerazione anche i vari approcci alla cognizione analizzati nel volume, per arrivare ad affrontare nello specifico le tesi che Dumouchel e Damiano propongono in relazione allo studio delle emozioni nell'ambito dell'artificiale e nel più ampio contesto dei vari tratti rilevanti della robotica sociale in generale, tra i quali trova posto anche l'aspetto etico.

2. Il ruolo epistemico dei robot sociali

L'idea secondo cui gli sviluppi della robotica sociale possono contribuire in modo significativo alla comprensione della mente, delle emozioni, della socialità umana è uno dei fili rossi che legano i vari capitoli di *Vivere con i robot*. È un'idea che gli autori esplorano lungo almeno due direttrici. La prima riguarda il ruolo sperimentale che i robot sociali possono svolgere nella scoperta dei meccanismi che regolano il comportamento sociale negli esseri umani. La seconda riguarda il ruolo che lo sviluppo della robotica sociale può svolgere nella comprensione della natura della mente e delle emozioni (per esempio, nello sviluppo della tesi secondo cui la mente “è la relazione tra agenti epistemici”, p. 102). Per brevità potremmo dire che, secondo gli autori, i robot sociali possono contribuire ad affrontare questioni *empiriche* (sui meccanismi del comportamento sociale) e questioni *ontologiche* (sulla natura della mente e delle emozioni). Vi sono ottimi

argomenti per aderire con convinzione alla prima tesi; la seconda – come dovremmo aspettarci, vista la sua originalità e la novità delle tecnologie a cui si fa riferimento – è in parte controversa e bisognosa di ulteriori giustificazioni.

Iniziamo dalla prima tesi. Sin dalla nascita della scienza moderna, la costruzione di automi ha contribuito in vario modo al progressivo consolidamento di concezioni meccanicistiche sul vivente (Riskin 2016; Truitt 2015). Tuttavia, è dall’inizio del XX secolo (Cordeschi 2002) che i robot vengono utilizzati come strumenti genuinamente *sperimentali* per valutare la plausibilità di ipotesi sui meccanismi del comportamento umano e animale, onorando il principio baconiano espresso dallo slogan “comprendere costruendo”. Uno dei modi in cui tale principio si è finora articolato è il cosiddetto “metodo sintetico”, discusso a più riprese da Dumouchel e Damiano. Il metodo sintetico prevede l’implementazione su un robot di un’ipotesi sul meccanismo che produce certi aspetti del comportamento di un sistema vivente, e il confronto tra il comportamento del robot e del sistema vivente in esame in un certo ambiente sperimentale. Se i comportamenti concordano, entro un certo livello di approssimazione e secondo particolari criteri di confronto, si può concludere che il meccanismo implementato nel robot costituisce un’ipotesi plausibile sui meccanismi che regolano gli aspetti del comportamento del sistema vivente oggetto di studio; altrimenti, l’ipotesi è rigettata. Secondo Dumouchel e Damiano, versioni di questa idea sperimentale possono essere applicate anche allo studio del comportamento sociale, attraverso l’implementazione di quelli che essi chiamano “sostituti” robotici. L’inserimento di un robot in un contesto popolato da esseri umani può essere visto in alcuni casi come un esperimento sociale: analizzando il modo in cui la presenza del robot modula l’interazione tra gli esseri umani coinvolti è possibile ottenere indicazioni interessanti sui meccanismi che governano la socialità umana.

Questa proposta¹ è di notevole interesse metodologico ed è un merito di Dumouchel e Damiano quello di contestualizzarla nell’ambito di una riflessione di grande respiro sulle frontiere della robotica sociale. Ci limitiamo qui a proporre alcuni spunti di approfondimento. Lo slogan “comprendere costruendo” è, appunto, uno slogan. Uno dei ruoli che la filosofia della scienza può svolgere in relazione a questi temi è quello di indagare in profondità le complessità metodologiche che emergono nelle applicazioni del metodo sintetico, le condizioni della sua stessa applicabilità, la forma delle “why-questions” che esso permette di

¹ Su cui si veda anche Wykowska, Chaminade & Cheng (2016).

affrontare. Analisi specificamente dedicate a questi temi (Webb 2000, 2001; Datteri & Tamburrini 2007) incoraggiano a interpretare con grande cautela lo slogan sopra citato. Si noti, peraltro, che la strategia sperimentale delineata da Dumouchel e Damiano non corrisponde effettivamente al metodo sintetico. Nel metodo sintetico si costruisce un robot e si analizza il suo comportamento per capire se il meccanismo che controlla il robot può essere considerato simile, sotto qualche aspetto e in qualche senso del termine, al meccanismo che controlla il sistema vivente in esame. Nella procedura delineata dai due autori, l'oggetto di analisi non è tanto il comportamento del robot quanto il comportamento che gli esseri umani producono in reazione alla presenza del robot. L'obiettivo, inoltre, non è quello di capire se il meccanismo che controlla il robot è simile (in qualche senso del termine) al meccanismo che governa il comportamento oggetto di studio: l'interesse è rivolto verso la scoperta dei meccanismi di interazione tra esseri umani. Si tratta sempre, in un certo senso, di "comprendere costruendo", ma le differenze procedurali sono sufficientemente marcate da suggerire la necessità di inaugurare analisi metodologiche specificamente dedicate alle complessità della strategia delineata da Dumouchel e Damiano. Per enfatizzare ulteriormente la distinzione, si noti per esempio che il metodo sintetico richiede che il robot implementi un'ipotesi sul meccanismo che determina il comportamento del sistema target. Questo vincolo, nelle applicazioni epistemiche della robotica sociale, può essere rilassato. Anche un robot del tutto discordante da qualsiasi ipotesi plausibile sui meccanismi del comportamento umano può essere utilmente inserito in una comunità di esseri umani per studiare, per esempio, i fattori che determinano l'emergere di "malfunzionamenti" sociali².

Ecco un'altra questione sollevata da Dumouchel e Damiano che merita approfondimento. Gli autori si allineano con la maggior parte degli odierni sostenitori del metodo sintetico nel ritenere che l'*embodiment* per definizione caratteristico dei robot fornisca una chiave di volta per affrontare con successo questioni relative al ruolo del corpo e dell'ambiente nella produzione del comportamento (sociale). Questa posizione non è ovvia. È certamente vero che il comportamento di una simulazione al calcolatore non è soggetto all'influsso di fattori ambientali (a parte eventi che danneggino l'hardware del calcolatore), che l'ambiente in cui vive l'organismo simulato debba essere anch'esso simulato (dunque largamente soggetto a processi di astrazione e idealizzazione), che una simulazione al calcolatore, a meno che non sia appunto robotica, non può trasformare

² Una riflessione metodologica sulla robotica interattiva è fornita in Datteri (2020).

direttamente il proprio ambiente circostante. Tuttavia, parlare del ruolo sperimentale dei robot (sociali) significa parlare del loro ruolo nella scoperta e nel controllo empirico di *teorie scientifiche* sui meccanismi del comportamento (sociale). Le teorie, quando sono ben definite, incorporano assunzioni piuttosto precise sulle caratteristiche fisiche e ambientali che vengono supposte rilevanti rispetto alla produzione del comportamento da spiegare, come ben distinte da quelle che invece la teoria ritiene irrilevanti. Le simulazioni al computer offrono la possibilità di un'implementazione "pulita" di tutti e solo quegli aspetti rilevanti, possibilità invece non così ovvia quando l'implementazione è robotica – il comportamento del robot, che è un oggetto fisico immerso in un ambiente fisico, sarà verosimilmente soggetto a miriadi di condizioni al contorno che non sono affatto menzionate nella teoria da controllare. Le simulazioni al computer, dunque, possono in alcuni casi – *modulo* le considerazioni ampiamente discusse nella filosofia delle simulazioni, si veda per esempio (Weisberg 2013; Winsberg 2010) – costituire strumenti per il controllo di teorie scientifiche relativamente più "trasparenti" rispetto ai robot. Anche di teorie scientifiche che diano peso al ruolo dell'ambiente e del corpo. Il fatto che le simulazioni al computer con buona probabilità falliranno nel riprodurre la complessità del comportamento (sociale) dei sistemi viventi non costituisce un buon argomento contro la loro validità come strumenti sperimentali: il fallimento di una simulazione nel riprodurre il comportamento del sistema in esame conduce al rifiuto di tale ipotesi, e il rifiuto di un'ipotesi scientifica contribuisce al progresso della scienza tanto quanto la sua provvisoria accettazione. Sotto quali condizioni, allora, gli esperimenti robotici sono "da preferirsi" rispetto alle simulazioni al computer per lo studio dei meccanismi che regolano il comportamento dei sistemi viventi?

Il secondo senso in cui Dumouchel e Damiano ritengono che i robot possano contribuire alla comprensione della mente, delle emozioni e del comportamento sociale umano verte sul loro ruolo nel dirimere questioni ontologiche sulla natura della mente e delle emozioni. La robotica sociale, affermano gli autori "non si limita a prendere posizione nel dibattito sulla natura della mente, ma ne trasforma i termini" (p. 61) in modi che verranno discussi nella sezione successiva. Ciò che gli autori intendono sostenere è che il comportamento dei robot sociali attuali o futuribili, oppure la possibilità di certe dinamiche di interazione tra robot ed esseri umani (per esempio, dinamiche in cui esseri umani provano empatia nei confronti dei robot; in cui robot reagiscono appropriatamente alle emozioni umane; in cui esseri umani sono disposti a delegare certi compiti a sostituti robotici),

possono aiutarci a capire *cos'è la mente e cosa sono le emozioni*. Fornendo, per esempio, buoni motivi per ritenere che “la mente si delinea non più come una macchina computante più o meno (dis-)incorporata, ma come una struttura di interconnessione – una rete cognitiva socialmente distribuita” (pp. 91-92). In queste righe non è a tema la plausibilità intrinseca delle tesi ontologiche difese dagli autori. La questione è a monte: in che senso il comportamento di un robot, o il dispiegarsi di una dinamica di interazione umani-robot, possono fornire buone ragioni per sostenere particolari tesi ontologiche sulla natura della mente e delle emozioni? Il comportamento individuale e sociale dei robot richiede una spiegazione, che fa necessariamente riferimento a certi enti teorici – appartenenti magari al dominio della mente e delle emozioni. Ma non è chiaro in che senso quelle spiegazioni implicano l'adozione di particolari tesi ontologiche sulla natura degli enti di cui esse “parlano”. Per esempio, Dumouchel e Damiano sottolineano con legittima forza il fatto che una buona interazione tra robot ed esseri umani possa essere spiegata soltanto assumendo che si verifichino particolari meccanismi di interazione mentale e affettiva tra i soggetti in gioco. Ma la scelta ontologica in merito a come interpretare le nozioni di “mente” e di “affetto”, in questa cornice teorica, è *a monte* e largamente sotto-determinata dal successo di queste spiegazioni. Si può dire che lo stesso vale, naturalmente, in quella che consideriamo interpretazione classica delle scienze cognitive. L'attribuzione di una mente alle macchine, per esempio, è in linea di principio compatibile con molte posizioni ontologiche sulla natura della mente – con posizioni dualistiche, con posizioni monistiche che affermano identità di occorrenza tra stati mentali e stati fisici, e persino, con alcuni accorgimenti, con tesi ispirate all'identità di tipo. *Vivere con i robot* solleva dunque una questione, di pertinenza dei filosofi della mente e della scienza, non chiarita a fondo nel testo stesso ma estremamente intrigante: anche accettando l'ipotesi (del tutto plausibile e ampiamente argomentata nel libro) secondo cui la robotica sociale può contribuire allo studio dei meccanismi mentali ed emotivi che regolano il comportamento sociale, in che senso tale disciplina può fornire nuovi e forti argomenti per dirimere questioni prettamente *ontologiche* sulla natura della mente e delle emozioni?

3. La filosofia della scienza cognitiva robotica e la questione dell'affettività

Analizzando più nello specifico la tesi in robotica affettiva proposta da Dumouchel e Damiano, si nota come essa sia supportata da una serie di considerazioni critiche rispetto ad alcuni filoni della filosofia della scienza cognitiva. L'argomentazione degli autori è originale nella misura in cui la trattazione dell'apparato teorico da un punto di vista filosofico viene introdotta a partire dall'etologia artificiale come "filosofia artificiale della mente animale" (p. 68). Tale filone di ricerca affonda le sue radici nella cibernetica e nello studio degli organismi biologici al fine di riprodurne i meccanismi di controllo e comunicazione, e dunque anche, e in particolare, negli organismi animali (Cordeschi, 2002) e ciò è in linea con il richiamo del metodo sintetico, da parte di Dumouchel e Damiano, la cui valenza abbiamo già discusso nel paragrafo precedente. Alla cibernetica, campo di studi costitutivamente interdisciplinare e a cui possiamo far risalire i momenti salienti del periodo precedente la nascita non solo dell'intelligenza artificiale in generale, ma anche, nello specifico, della robotica e di altri sottocampi delle scienze dell'artificiale come l'*artificial life*, gli autori si richiamano solo in parte. La scelta di chiamare in causa l'etologia artificiale è proattiva alla discussione degli approcci cognitivi che costituiscono il quadro in cui inserire la tesi proposta da Dumouchel e Damiano, ovvero l'*embodied cognition* nelle sue forme più radicali (Chemero 2009).

Il passaggio verso questo approccio alla cognizione è al tempo stesso motivato e sviluppato da una critica ad approcci più tradizionali, fondati su una tradizione di studio degli animali come macchine che si può far risalire fino a Descartes e che ha assunto nella contemporaneità come principio base una forma di dualismo di carattere non ontologico. La critica del dualismo cartesiano non è di per sé una critica a Descartes in sé, ma al modo in cui il suo pensiero è stato recepito e trasformato nel tempo, e gli autori sottolineano, contro una tradizione ormai superata, il ruolo di Descartes nel riconoscere agli animali funzioni cognitive pur nella simultanea negazione di una mente animale (pp. 72-73). Una tale prospettiva ha il fine di vedere in Descartes un propugnatore del pluralismo cognitivo: sistemi diversi, quali sono gli animali e gli esseri umani, esibiscono sì tutti capacità cognitive, ma in modo differente. Se questa interpretazione costituisce una riabilitazione di Descartes, che trova certamente supporto nei diversi passi *Discorso sul metodo* in cui Descartes parla delle facoltà cognitive che distinguono gli

esseri umani dagli animali, tra cui fra tutte il linguaggio³, allo stesso tempo ne fa per Dumouchel e Damiano il propugnatore di un dualismo più sottile, dall'esito paradossale: la mente diventa il tratto distintivo di un sistema cognitivo tra altri sistemi cognitivi, che per tale qualità risulta essere quello più raffinato, complesso, inattuabile. Una simile interpretazione del dominio del mentale difficilmente può essere condivisa da chi ritiene che un sistema cognitivo è anche un sistema, in un qualche modo, mentale. Questo vale specialmente per coloro che ritengono che il mentale si possa considerare, in un quadro epistemologico di ricerca su di esso, come cognitivo nel senso di essere un sistema che funziona elaborando informazioni. Gli autori, tuttavia, sposando un approccio diverso, meno "classico" alla cognizione, vedono nel dualismo sopra descritto, basato su una differenza "di grado e non di natura" (p. 75), un tipo di dualismo che caratterizza tante posizioni contemporanee in filosofia della mente e delle scienze cognitive, anche se non sempre è riconosciuto – e a volte addirittura è negato – dagli specialisti che sostengono tali posizioni.

Queste basi motivano il rigetto deciso di non pochi approcci contemporanei alla cognizione, fra cui pure alcuni in contrasto con (o che almeno intendono costituire un superamento de) la cosiddetta scienza cognitiva classica, in larga parte assimilabile nei principi e negli assunti all'intelligenza artificiale simbolica⁴. In tal senso va vista la critica della visione della mente estesa, che pure sembrerebbe un buon candidato a supporto di tutta la "tecnologia intelligente" *embodied*, robotica compresa. La mente estesa, per gli autori, non è altro che un'estensione del dualismo nel senso sopra esposto, in cui la mente diviene egemone nel dominio del cognitivo fagocitando e trasformando in qualcosa di omogeneo a sé tutto ciò che entra in contatto *attivamente* con essa (p. 80). La critica alla mente estesa e all'idea di cyborg "metafisici", "intellettuali", cioè disincorporati (p. 78) è in linea con la letteratura sul dibattito filosofico, che riconosce l'impianto ancora funzionalista di questa ipotesi, fortemente debitore dell'assunzione della realizzabilità multipla del mentale⁵, seppure gli autori ne vedano le debolezze maggiori nella mancata caratterizzazione delle dimensioni dinamiche del mentale e nella relega della mente alla sua individualità solipsistica.

³ Si veda ad esempio la parte V del *Discorso sul metodo*.

⁴ Per una ricostruzione storica dettagliata di questi aspetti e in generale dell'evoluzione della scienza cognitiva fino agli inizi del ventesimo secolo si veda Boden (2006).

⁵ Si vedano ad esempio le osservazioni di Marraffa & Paternoster (2012, 138-139).

Lo stesso tipo di critica, non sorprendentemente, viene rivolta anche alla nozione di “marchio del cognitivo” nel modo in cui, ad esempio, è elaborato da Rowlands (2010), secondo cui un sistema è cognitivo se gli viene resa disponibile come soggetto intenzionale informazione di cui prima non è in possesso, sia che essa provenga dai suoi “meccanismi” interiori, subpersonali, sia da elaboratori esterni (p. 86-87). Ciò dovrebbe valere anche per un sistema robotico che possa dirsi cognitivo. Anche la prospettiva del marchio del cognitivo è, secondo gli autori, troppo antropocentrica ed è ragionevole supporre che ogni allargamento della prospettiva a una visione più onnicomprensiva dal punto di vista intenzionale, verso approcci che difendano un più essenziale “marchio del mentale”, vada incontro allo stesso tipo di critiche da parte di Dumouchel e Damiano. Va tuttavia sottolineato che il marchio del cognitivo è di per sé una nozione introdotta al fine di portare una critica all’ipotesi della mente estesa e usata soprattutto per una difesa del carattere intracranico della cognizione⁶. In direzione opposta alle critiche rivolte dagli autori a tali approcci, da loro considerati ancora cartesiani, alla scienza cognitiva ci si potrebbe chiedere in che modo gli autori potrebbero definire un contrassegno del robotico, una sorta di “*mark of the robotic*”. Al di là dei concetti di autonomia, autorità e sostituzione, può la robotica dotarsi di un carattere essenziale? Avrebbe senso uno sforzo definizionale in questa direzione che tenga conto delle accuse di metafisicità rispetto ad agenti cognitivi di questo tipo che ne superi il tratto solipsistico? Può darsi un marchio del robotico che includa, in questo orizzonte di analisi della cognizione, gli aspetti che appaiono ineliminabili nel definire un artefatto robotico anche come agente cognitivo, ovvero *l’embodiment*, l’automazione e l’autonomia nelle scelte sia dell’azione che dei metodi di decisione?

La domanda, in generale, è aperta, ma gli spunti verso questa direzione non mancano in *Vivere con i robot*. Ad esempio, in tal senso appare rilevante la questione del pluralismo. Al pluralismo dei sistemi cognitivi secondo gli autori va sostituito un altro tipo di pluralità, quella che vede coinvolti più attori di un processo che ha come risultato l’emergenza del mentale (p. 76). Su questa linea va visto il richiamo della posizione di *radical embodiment* in cui la teoria dell’affettività robotica viene sviluppata dagli autori, cioè *l’enaction* (p. 68). L’enattivismo come approccio di *embodiment* radicale alla cognizione si rifà al pensiero di Maturana, Varela e Thompson (Maturana & Varela 1980; Thompson & Varela 2001) e ha

⁶ Adam & Aizawa (2010). Per una discussione e una difesa dell’ipotesi della mente estesa si veda Piredda (2017).

trovato un rinnovato interesse da parte della filosofia della scienza cognitiva anche per il recupero di istanze fenomenologiche all'interno degli studi sulla cognizione (Gallagher 2017). Nella prospettiva enattiva il sistema cognitivo non è più considerato nel suo isolamento, ma da un punto di vista totalmente opposto che lo vede incorporato e strettamente connesso all'ambiente in cui opera. Il confine tra sistema e ambiente non è un limite, bensì il punto di contatto che rende l'insieme il luogo in cui emergono comportamenti complessivi che spiegano la dinamica di coevoluzione alla base della specifica capacità cognitiva che si intende spiegare. In tal senso, si può parlare, nell'ambito della robotica sociale, di più attori coinvolti nel processo di emergenza e dunque di un rinnovato tipo di pluralità.

La tesi principale proposta in *Vivere con i robot* si inquadra all'interno di questo approccio cognitivo e aspira a spiegare il ruolo delle emozioni nella robotica sociale proponendo un punto di vista originale sulle emozioni, che possa portare all'identificazione di una sorta di empatia artificiale. Anche in merito allo specifico discorso sulle emozioni gli autori partono da una critica al cartesianesimo della scienza cognitiva contemporanea: la mente non è vista essere in un agente, un sistema o un cervello, bensì nella relazione stessa fra agenti epistemicici (p. 102). La mente come tale, in altri termini, emerge principalmente in un ambiente sociale. Tale è anche il destino delle emozioni, così come dello studio delle emozioni in robotica, che beneficiano dell'essere il risultato emergente di agenti robotici relazionali inseriti nella rete delle interazioni con altri robot o esseri umani. In tale quadro, la tesi originale di Dumouchel e Damiano consiste nel superamento degli approcci "esterno" ed "interno" allo studio delle emozioni in robotica (p.108). La "robotica esterna" si caratterizza come lo studio della simulazione dei tratti emozionali umani e animali in un robot. In tale prospettiva le emozioni sono dunque simulate, "finte". La "robotica interna" si occupa invece di come riprodurre nell'artefatto artificiale i meccanismi alla base delle emozioni negli esseri viventi, considerando uno dei propri fini quello che le emozioni siano autentiche. Questo secondo caso è quello più affine al metodo sintetico che abbiamo delineato in precedenza, eppure per gli autori entrambi sono viziati da un dualismo di fondo, poiché la realtà delle emozioni è comunque relegata all'interno di un agente, un artefatto robotico in questo caso, e tutto ciò che è esterno è solo inautenticità o conseguenza passiva di ciò che avviene dentro il sistema cognitivo stesso, nelle parti e nelle connessioni della sua architettura.

La tesi che gli autori propongono vuole oltrepassare da una parte, gli aspetti antropomorfici della robotica sociale esterna già richiamati

dall'ipotesi dell'*uncanny valley* e l'idea che le emozioni siano qualcosa di evocato *dai* robot e attribuito dall'esterno *ai* robot; dall'altra, il ritorno a una visione internista della cognizione che manca di cogliere, come già detto, gli aspetti dinamici e coevolutivi della cognizione trattabili e spiegabili solo grazie alla comprensione dell'interazione e della coevoluzione sistema/ambiente. Oltre questa dicotomia, gli autori analizzano anche la prospettiva interazionista che supera l'opposizione internismo/esternismo in merito alle emozioni robotiche. I robot diventano attori sociali in una dinamica che coinvolge progressivamente l'utente in una sorta di circolarità che si autorafforza (p. 128). In questa prospettiva, cade il vincolo di riprodurre un'architettura interna che sia l'implementazione di plausibili meccanismi emozionali biologici, perché non è rilevante nell'interazione utente/agente artificiale. Ciò che conta è la loro interazione e l'emergenza di una risposta di coinvolgimento affettivo sempre più elevata. In altri termini, il robot non conta per come è costruito, il che costituisce peraltro una violazione del metodo sintetico classico, ma per come agisce in risposta all'utente nella dinamica di interazione affettiva. Tuttavia, cade inoltre anche un altro tratto saliente, questo dell'approccio esterno, della robotica affettiva condotta in termini dicotomici: il carattere simulato delle emozioni robotiche:

dal punto di vista degli specialisti, le emozioni artificiali, configurate in tal modo [prescindendo cioè dalla costruzione ispirata da meccanismi naturali], non possono più essere interpretate come mezzi per "ingannarci", portandoci a pensare che i robot "abbiano" un'emozione quando così non è. Si pensa che la correlazione delle espressioni emozionali alla dinamica interna degli agenti robotici renda le manifestazioni affettive di queste macchine più "vere" – più genuine. (pp. 131-132)

Anche la prospettiva interattiva viene rigettata dagli autori, perché non coglie l'obiettivo di superare la dicotomia interno/esterno, pure nel tentativo di ridefinirla in una prospettiva integrata. La proposta di Dumouchel e Damiano è di abbandonare completamente tale dicotomia, anche là dove essa è caratterizzata come interazione, in vista di una visione radicale che trasformi l'interazione in un approccio totalmente intersoggettivo, tipico dell'enazione. Non più di interazione dunque si deve parlare ma di coordinazione fra agenti, una coordinazione in cui gli agenti si influenzano reciprocamente in un processo dinamico e mai concluso, se non con la cessazione della relazione sociale. Estendere la prospettiva enattiva alla modellizzazione delle emozioni robotiche significa portare in primo piano, anzi su un piano esclusivo, le dinamiche relazionali e sociali che fanno

emergere gli aspetti emozionali fra agenti che coevolvono influenzandosi a vicenda. Restano però aperte alcune domande. Tale visione radicale permetterà una caratterizzazione effettiva delle emozioni robotiche in relazione all'utente? Essa costituisce un reale superamento di ogni paradigma dicotomico-oppositivo di spiegazione delle emozioni in ambito robotico? Se da un punto di vista esplicativo le teorie e le modellizzazioni che ne derivano sembrano beneficiare del rifiuto radicale di ogni opposizione dualistica nello studio della cognizione, tra cui quella di antica matrice filosofica fra soggetto e oggetto, la domanda ontologica sulle emozioni dal punto di vista robotico sembra restare irrisolta. Come anche esplicitato nel passo precedentemente riportato, la richiesta di emozioni più vere, più genuine, sembra essere una delle richieste principali della robotica affettiva. Tuttavia, nulla garantisce che essa non sia fuori dei limiti anche della spiegazione enattiva. Il problema ontologico di cosa siano le emozioni resta inespresso e questo, se non si vuole vedere come un ostacolo di tipo metafisico, perché attinente a riflessioni che appartengono ad altri contesti di ricerca, ha una rilevanza di tipo epistemologico nelle teorie che vogliono affrontare, sinteticamente o meno, lo studio delle emozioni e dell'empatia nell'ambito dell'artificiale.

4. Conclusione: un programma di ricerca

La filosofia della robotica è una filosofia della scienza che chiama in causa più aspetti di quelli generalmente trattati dalla filosofia della scienza generale, non solo perché si tratta di una filosofia di una scienza speciale, ma anche perché deve confrontarsi con diversi ambiti disciplinari, da quelli legati alla cognizione a quelli più tecnologici, fino ad abbracciare tematiche eminentemente appartenenti ad altre branche della filosofia, come l'etica, la riflessione politica o l'estetica, in particolare per quanto riguarda la percezione. Tutto ciò è particolarmente evidente nel libro di Dumouchel e Damiano. Un'argomentazione supportata da una scrittura varia e complessa, ma sempre chiara, rende il volume ricco di spunti. Proprio a causa della ricchezza delle tematiche affrontate, non sono del tutto assenti alcune mancanze. Ad esempio, le riflessioni legate all'etica sintetica non sembrano avere l'adeguata trattazione che potrebbero avere all'interno del quadro tracciato per affrontare gli aspetti della robotica affettiva, e i due sensi di etica robotica che vengono esposti, pur nel recupero delle nozioni di autonomia e di "sostituto", non chiariscono quanto si potrebbe volere il

modo in cui tale trattazione dell'etica robotica possa beneficiare di un approccio coordinativo e sistemico.

Nonostante questo, l'accostamento tra robotica affettiva ed etica sintetica presente nel libro ha il beneficio di consentire un allargamento della prospettiva proposta in relazione alla tesi principale del libro verso altri aspetti della robotica sociale, generando ulteriori spunti di dibattito che certamente troveranno svolgimento in futuro. Questo non tanto in relazione alle sfide etiche poste dall'intelligenza artificiale, quanto bensì in riferimento a come dotare, e al senso che può avere, un artefatto robotico di principi comportamentali o schemi di comportamento con un grado maggiore o minore di autonomia nell'interazione con altri sistemi intelligenti ed "etici". Ad esempio, un approccio sintetico all'etica in una prospettiva coordinativa potrà gettare ulteriore luce sulla situazione apparentemente paradossale per cui dotare un sistema autonomo di regole non è renderlo un sistema a tutti gli effetti etico se il sistema stesso è *costretto* a seguire le regole. In altri termini, si tratta di uno degli aspetti più generali del problema dell'effettivo decisore per quanto riguarda il comportamento di un sistema etico, già ampiamente posto e affrontato nella letteratura di riferimento (si vedano, ad esempio, Moor, 2006; Anderson & Anderson, 2011).

Le conclusioni cui giungono Dumouchel e Damiano propongono un programma di ricerca unitario, che, nella prospettiva dell'approccio coordinativo ed enattivo tra artefatto robotico e agente umano, definisca le possibilità e i limiti di una sperimentazione volta a cogliere gli aspetti emergenti del comportamento complessivo, sistemico, di agenti naturali e artificiali (robotici) – pur con i vincoli sempre più sfumati di questa distinzione – che a loro volta, singolarmente, mettono in atto un comportamento guidato da regole. Se la spinta verso una rimodulazione del modo in cui vengono studiate le emozioni artificiali in robotica sociale appare giustificata e foriera di spunti di riflessione, restano aperte varie domande come, ad esempio, quelle sui confini applicativi del metodo sintetico, sulle effettive risultanze di una sperimentazione che utilizzi tale metodologia, con i vincoli epistemologici cui è connessa, in campi come quello affettivo, politico o morale nell'abito dell'artificiale. Ciò che sembra certo, o quantomeno difficilmente dubitabile – e l'averlo sottolineato è uno dei meriti del volume *Vivere con i robot* – è che una continua ridefinizione degli standard e degli approcci sperimentali all'interazione essere umano/robot non solo appare inevitabile, ma si accompagna anche inevitabilmente alle presenti e future discussioni filosofiche ed

epistemologiche relative agli approcci alla cognizione che sottendono lo sviluppo delle applicazioni robotiche umanoidi, di sostituti e di sistemi autonomi in generale.

Bibliografia

Adams F., Aizawa K., 2010, «Defending the bounds of cognition», in Menary R. (ed), *The Extended Mind*, Cambridge, Mass., MIT Press, pp. 67-80.

Anderson M. Anderson S. L., (eds.) (2011), *Machine Ethics*, Cambridge , Cambridge University Press.

Boden M., 2006, *Mind as Machine. A History of Cognitive Science*, Oxford University Press, Oxford.

Breazeal C., Dautenhahn K., Kanda T., 2016, «Social Robotics», in Siciliano B., Khatib O. (eds.), *Springer Handbook of Robotics*, Cham, Springer, pp. 1935-1972.

Chemero A., 2009, *Radical Embodied Cognitive Science*, Cambridge, Mass., MIT Press,

Cordeschi R., 2002, *The Discovery of the Artificial. Behavior, Mind and Machines Before and Beyond Cybernetics*, Dordrecht, Kluwer Academic Publishers.

Datteri E., 2020, «Interactive biorobotics», *Synthese*, Online Fir. <https://doi.org/10.1007/s11229-020-02533-2>

Datteri E., Tamburrini G., 2007, «Biorobotic Experiments for the Discovery of Biological Mechanisms», *Philosophy of Science*, 74, 3, pp. 409-430.

Gallagher S., 2017, *Enactivist interventions. Rethinking the Mind*, Oxford, Oxford University Press.

Marruffa M., Paternoster A., 2012, *Persone, menti, cervelli. Storia, metodi e modelli delle scienze della mente*, Milano, Mondadori.

- Maturana H. R., Varela F. J., 1980, *Autopoiesis and Cognition. The Realization of the Living*, Dordrecht, Reidel.
- Moor, J. (2006), «The Nature, Importance, and Difficulty of Machine Ethics», *IEEE Intelligent Systems*, 21, 4, pp. 18-21.
- Mori M., 1970, «Bukimi no tani» («The uncanny valley»), *Energy*, 7, 4, pp. 33-35 (translated and reprinted by K. F. MacDorman and N. Kageki in *IEEE Robotics & Automation Magazine*, 19, 2, pp. 98-100).
- Piredda G., 2017, «The Mark of the Cognitive and the Coupling-Constitution Fallacy: A Defense of the Extended Mind Hypothesis», *Frontiers in Psychology*, 8, 2061.
- Riskin J., 2016, *The Restless Clock. A History of the Centuries-Long Argument over What Makes Living Things Tick*, Chicago, University of Chicago Press.
- Rowlands M., 2010, *The New Science of the Mind. From Extended Mind to Embodied Phenomenology*, Cambridge, Mass, MIT Press.
- Thompson E., Varela F. J., 2001, «Radical embodiment. Neural dynamics and consciousness», *Trends in Cognitive Sciences*, 5, 10, pp. 418-425.
- Truitt E. R., 2015, *Medieval Robots. Mechanism, Magic, Nature, and Art*, Philadelphia, University of Pennsylvania Press.
- Webb B., 2000, «What does robotics offer animal behaviour?», *Animal Behaviour*, 60, 5, pp. 545-558.
- Webb B., 2001, «Can robots make good models of biological behaviour?», *Behavioral and Brain Sciences*, 24, pp. 1033-1050.
- Weisberg M., 2013, *Simulation and Similarity: Using Models to Understand the World*, Oxford, Oxford University Press.
- Winsberg E. B., 2010, *Science in the Age of Computer Simulation*, Chicago, University of Chicago Press.

Wykowska A., Chaminade T., Cheng, G., 2016, «Embodied artificial agents for understanding human social cognition», *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 1693.

APhEx.it è un periodico elettronico, registrazione n° ISSN 2036-9972. Il copyright degli articoli è libero. Chiunque può riprodurli. Unica condizione: mettere in evidenza che il testo riprodotto è tratto da www.aphex.it

Condizioni per riprodurre i materiali --> Tutti i materiali, i dati e le informazioni pubblicati all'interno di questo sito web sono "no copyright", nel senso che possono essere riprodotti, modificati, distribuiti, trasmessi, ripubblicati o in altro modo utilizzati, in tutto o in parte, senza il preventivo consenso di APhEx.it, a condizione che tali utilizzazioni avvengano per finalità di uso personale, studio, ricerca o comunque non commerciali e che sia citata la fonte attraverso la seguente dicitura, impressa in caratteri ben visibili: "www.aphex.it". Ove i materiali, dati o informazioni siano utilizzati in forma digitale, la citazione della fonte dovrà essere effettuata in modo da consentire un collegamento ipertestuale (link) alla home page www.aphex.it o alla pagina dalla quale i materiali, dati o informazioni sono tratti. In ogni caso, dell'avvenuta riproduzione, in forma analogica o digitale, dei materiali tratti da www.aphex.it dovrà essere data tempestiva comunicazione al seguente indirizzo (redazione@aphex.it), allegando, laddove possibile, copia elettronica dell'articolo in cui i materiali sono stati riprodotti.

In caso di citazione su materiale cartaceo è possibile citare il materiale pubblicato su APhEx.it come una rivista cartacea, indicando il numero in cui è stato pubblicato l'articolo e l'anno di pubblicazione riportato anche nell'intestazione del pdf. Esempio: Autore, *Titolo*, <<www.aphex.it>>, 1 (2010).
