
GIOVANNI SARTOR

GLI AGENTI SOFTWARE E LA DISCIPLINA GIURIDICA DEGLI STRUMENTI COGNITIVI

SOMMARIO: 1. Introduzione. Premessa metodologica. — 2. Una caratterizzazione degli AS. — 3. Gli stati cognitivi nelle entità artificiali. — 4. AS e responsabilità civile. — 5. Agenti e contratti. — 6. Conclusione.

1. INTRODUZIONE. PREMESA METODOLOGICA.

Come è noto, gli agenti software (AS) sono sistemi informatici in grado di agire con autonomia, senza il controllo diretto del loro utilizzatore. La tecnologia degli AS trova impiego nelle aree più diverse dell'informatica: nello sviluppo di sistemi informativi, nel co-ordinamento di applicazioni distribuite, nella realizzazione d'interfacce, nella gestione di attività complesse e di situazioni critiche (ad esempio, nel governo del traffico marittimo o aeroportuale). Per il giurista presentano un particolare interesse i casi in cui gli AS non esauriscono la propria attività nella sfera del loro utilizzatore (come avviene nell'uso di AS nella gestione decentrata della logistica aziendale), ma entrano in contatto con altri soggetti. Ciò accade soprattutto nella rete (Internet), dove operatori e utenti ricorrono agli AS per cogliere e sfruttare l'enorme varietà di possibilità offerte dallo sconfinato, diversificato e dinamico ambiente virtuale del cibernazio. Si tratta, in particolare, dell'uso degli AS nella raccolta di informazioni (ricerca di dati, di offerte relative a beni e servizi, di possibili partner commerciali) e nell'attività contrattuale (trattativa, formulazione di proposte, valutazione della loro convenienza, conclusione di contratti).

Gli AS sono una nuova realtà, spesso ancora allo stadio sperimentale (benché siano già disponibili numerose applicazioni commerciali), e in rapida evoluzione. Quindi non possiamo stabilire

* Il presente lavoro, realizzato nell'ambito del progetto europeo IST FET UIE ALFEBITE (IST-1999-10298), è stato presentato al convegno « Le nuove sfide del

diritto dell'informatica. Convegno di studi in ricordo di Ettore Giannantonio - Roma venerdì 6 dicembre 2002 ».

con precisione quali attività, funzioni, ruoli gli AS, svolgeranno in futuro nel commercio elettronico e nelle società elettroniche. I limiti delle nostre capacità di previsione, tuttavia, non debbono trattenerci dallo studiare come gli AS siano regolati dal diritto vigente e come dovrebbero essere regolati dal diritto futuro. Dobbiamo farlo ora, sia perché dobbiamo già oggi affrontare alcuni problemi giuridici connessi con gli AS, sia perché la consapevolezza dei problemi giuridici e delle possibili soluzioni può migliorare la progettazione delle future architetture tecnologiche, architetture che oggi si stanno realizzando a livello prototipale. Infine, dobbiamo esaminare gli AS da una prospettiva giuridica poiché ciò può suggerirci nuove idee per affrontare problemi generali atinenti al diritto dell'informatica. Gli AS, infatti, evidenziano un aspetto presente, seppure in modo latente, anche in altre applicazioni informatiche, ma al quale finora la dottrina non ha dedicato specifiche riflessioni: la delegazione di funzioni cognitive a strumenti automatici.

Lo studio dei problemi giuridici degli AS richiede una metodologia appropriata, che rifletta la novità dell'oggetto che stiamo esaminando. Si osservi che frequentemente tendiamo ad accostarci a nuove problematiche seguendo un procedimento di tipo classificatorio: per scoprire come il diritto regoli un certo oggetto, cerchiamo l'etichetta giusta (il concetto o la classe sotto la quale ricade quell'oggetto), apponiamo l'etichetta all'oggetto (lo classifichiamo o « sussumiamo » nel concetto), e quindi applichiamo la disciplina che il diritto associa a tale etichetta.

Secondo questo orientamento, per scoprire la disciplina giuridica degli AS, dovremmo classificare gli stessi in categorie giuridiche preesistenti. Dovremmo quindi porci questioni come le seguenti:

- Gli AS sono soggetti o oggetti?
- Gli AS sono persone o strumenti?
- Gli AS sono individui (moralmente o giuridicamente) responsabili, o sono solo cose?

La presupposizione dell'orientamento classificatorio (o concettualista, come anche si dice) è che la disciplina giuridica riguardante una certa entità segua automaticamente, una volta che abbiamo inserito l'oggetto nella classe cui appartiene. La classificazione è facilitata dal fatto che le categorie giuridiche si presentano spesso in dicotomie esaustive: il primo termine di ciascuna di esse è circoscritto positivamente, mentre il secondo è definito negativamente. Conseguentemente, l'incapacità di dimostrare che un'entità appartiene al primo termine della dicotomia mostra che esso appartiene al secondo. Così se non si riesce a stabilire che un'entità è una persona, allora si dovrà concludere che si tratta di uno strumento; se non si riesce a stabilire che un'entità è un soggetto, allora si dovrà concludere che essa è un oggetto; se non si riesce a

stabilire che un'entità è un individuo, allora si dovrà ritenere che essa è una cosa.

Inoltre, una volta che si sia stabilito che un'entità appartiene ad una certa classe, si tenderà ad applicare ad essa tutte le qualificazioni giuridiche associate a quella classe, anche quando tali qualificazioni siano giustificate da caratteristiche che non riguardano l'entità in questione. Così, chi assume che gli AS, e più in generale i sistemi informatici siano oggetti, e ritiene che gli oggetti in generale non possano avere stati cognitivi, concluderà apoditticamente che anche gli agenti e i sistemi informatici sono incapaci di assumere stati cognitivi rilevanti per il diritto.

Infine, la metodologia della classificazione può essere corretta mediante la tecnica delle finzioni: quando una certa classificazione conduca a risultati giuridici evidentemente assurdi, tali risultati si possono evitare assumendo falsamente che esistano circostanze che impediscono di derivare quei risultati. Così, per evitare la conclusione assurda che tutti i contratti conclusi da sistemi informatici senza diretto controllo umano sono invalidi (pur assumendo che ogni contratto richiede la volontà di una persona, e che un computer non è una persona), si può adottare la finzione che la persona nel cui interesse il contratto è stato concluso volesse concluderlo (anche se quella persona non era a conoscenza del fatto che quel particolare contratto veniva stipulato). La finzione ci consente quindi di mantenere inalterate tutte le nostre assunzioni e classificazioni, pur evitando le conclusioni giuridiche che ne seguono nel caso particolare.

In talune circostanze, il richiamo alla finzione può essere sostituito dal richiamo alla protezione dell'affidamento. La protezione dell'affidamento può giustificare la seguente equiparazione: l'erronea credenza che esista un certo stato di cose può avere gli stessi effetti giuridici che avrebbe l'esistenza di tale stato di cose. Ad esempio, nei contratti il fatto che una parte creda che la controparte abbia fatto una certa dichiarazione (che la controparte non ha fatto) o avesse una certa intenzione (che la controparte non aveva) può talvolta produrre lo stesso effetto che si sarebbe avuto qualora la controparte avesse fatto quella dichiarazione o avesse avuto quell'intenzione. Ciò avviene per proteggere la parte che era giustificata nel credere che la controparte avesse fatto tale dichiarazione o avesse tale intenzione (per proteggere la parte che aveva fatto affidamento in tale credenza). Tuttavia, non si può invocare la protezione dell'affidamento quale ragione per la quale si produrrebbero gli effetti di una certa situazione (pur nell'assenza di tale situazione) quando risulti che nessuno credeva che tale situazione esistesse, o quando era totalmente irragionevole credere che essa esistesse. Inoltre, di regola, l'erronea credenza che esista un certo stato di cose (mentale o fisico) può essere equiparata all'esistenza di tale stato di cose, solo quando la controparte abbia contribuito

in qualche modo a produrre tale erronea credenza, mediante le proprie azioni o omissioni.

Chi segue l'orientamento classificatorio, accostandosi agli AS è condotto ad affermare che essi non sono soggetti, non sono persone o individui. Da ciò egli tende ad inferire che gli AS sono solo oggetti, strumenti o cose, di modo che essi non possono avere una volontà indipendente. Infine, ciò lo conduce a concludere che gli AS non possono agire autonomamente, e non possono avere credenze, volontà o altri stati cognitivi giuridicamente rilevanti, di modo che risulta necessario attribuire agli utilizzatori degli AS ogni azione da questi compiuta, assumendo che gli utilizzatori conoscessero o dovessero conoscere, e quindi volessero o dovessero volere (secondo il principio della protezione dell'affidamento) quanto facevano i loro AS.

Questo sillogismo è in realtà fallace, come vedremo nel seguito. In particolare, l'idea che l'utilizzatore sappia esattamente che cosa farà il suo AS (o, in generale, il suo sistema informatico) è spesso solo una finzione. Questa conoscenza è impedita da quanto segue: *a*) l'ignoranza delle circostanze dei diversi contesti nei quali l'AS opererà; *b*) l'impossibilità di riprodurre, almeno per agenti sufficientemente complessi, tutte le operazioni di elaborazione che il software (dell'AS) eseguirà in ciascuna di tali circostanze; *c*) il fatto che l'utilizzatore non ha a disposizione tutte le informazioni che sono o saranno disponibili al suo AS; *d*) l'impossibilità di prevedere (almeno per AS che possano apprendere, e quindi modificare se stessi) quali dati e quali istruzioni faranno parte dell'AS quando esso opererà.

La falsità dell'assunto che l'utilizzatore sapesse che il suo AS avrebbe fatto quanto ha fatto, non può essere rimediata richiamandosi alla protezione dell'affidamento, quale giustificazione per assimilare la situazione nella quale l'utilizzatore non fosse a conoscenza del comportamento dell'AS, alla situazione nella quale egli sapesse e volesse l'azione di questo. In particolare, rispetto alla formazione di contratti, ciò condurrebbe a concludere che i contratti conclusi autonomamente da un AS, nel nome del suo ignaro utilizzatore, sono efficaci solo quando la controparte erroneamente (ma ragionevolmente) credesse che l'utilizzatore avesse preparato e voluto quei contratti. Quando la controparte non avesse questa credenza, non ci si può richiamare alla protezione dell'affidamento quale ragione affinché sia valido un contratto stipulato da un AS in nome e nell'interesse, ma senza la supervisione, del suo utilizzatore.

Si assuma, ad esempio, che il contraente riceva una dichiarazione proveniente da un AS in nome e nell'interesse dell'utilizzatore dell'AS stesso, e sia consapevole che l'utilizzatore non sa che l'AS sta facendo quella particolare dichiarazione: il contraente è a conoscenza del fatto che l'utilizzatore dell'AS non ha predisposto il contenuto del contratto, ma ha delegato all'AS l'ac-

certamento delle circostanze del caso, la determinazione conseguente del contenuto del contratto, e la deliberazione se concluderlo o meno. In questa situazione, il contraente sa che l'utilizzatore dell'AS non ha la specifica intenzione di stipulare quel particolare contratto (pur non avendo neppure l'intenzione di non stipularlo), e pertanto non esiste alcun affidamento (nell'esistenza di tale specifica intenzione) da proteggere¹.

Si consideri inoltre la situazione nella quale un contraente crede irragionevolmente che la dichiarazione contrattuale effettuata da un sistema informatico sia voluta dalla persona per conto della quale il computer sta agendo: la comprensione del contraente è così limitata che egli crede che, quando interagisce con Amazon.com, Mr. Bezos o uno dei suoi dipendenti sia seduto ad una tastiera all'altro capo del filo telefonico. È molto difficile immaginare perché il diritto dovrebbe prendere in considerazione una credenza tanto assurda. E come dire che il diritto dovrebbe prendere in considerazione la credenza che il commesso (umano) di un negozio quando conclude un contratto (ad esempio, decidendo di concedere uno sconto) stia solo trasmettendo la volontà del titolare del negozio (essendo in contatto telepatico o simile con quest'ultimo).

Pertanto l'idea della protezione dell'affidamento, pur potendo essere correttamente richiamata in alcune circostanze, non può fornire una via d'uscita al giurista che voglia ottenere risultati ragionevoli circa la disciplina degli AS (in modo da consentire a chi sia interessato di delegare alcuni compiti ad un AS), e conservare le classificazioni che gli sono care (gli AS sono solo oggetti, strumenti o cose) e le presunte implicazioni di tali classificazioni (gli AS non possono intraprendere azioni autonome, accompagnate da stati cognitivi corrispondenti). Se ci vogliamo allontanare dalle finzioni ed affrontare la realtà è necessaria una nuova impostazione. Ma per il pensiero classificatorio niente è mai realmente nuovo: ogni cosa può essere ricondotta ad una categoria preesistente.

La contestazione del metodo classificatorio nel diritto è molto risalente, e puntali critiche degli eccessi del concettualismo furono formulate già tra la fine del XIX secolo e l'inizio del XX, sia nei paesi di *civil law* sia in quelli di *common law* (in particolare, ricordiamo la *Interessenjurisprudenz* in Germania e la *Sociological Jurisprudence* negli Stati Uniti). Qui, non potendo ripercorrere le vicende del dibattito sul metodo giuridico, ci limitiamo ad osservare

¹ Stiamo ora considerando l'affidamento del contraente che interagisce con l'AS, e non la buona fede dello stesso AS, di modo che la discussione se si debba adot-

tare una nozione oggettiva o una nozione soggettiva di buona fede in relazione agli agenti non ha rilievo (cf. Weitzenboeck 2002).

che un'impostazione di tipo classificatorio può essere utile in certe circostanze, in quanto consente di trovare velocemente una soluzione giuridica *prima facie* (mediante classificazione) e poi di fare qualche aggiustamento mediante le opportune finzioni. Tuttavia, questo metodo è completamente inappropriato quando si voglia sviluppare un modello concettuale coerente, che consenta di accostarsi ad un nuovo fenomeno (come gli AS) e di formulare le ragioni a sostegno di nuove soluzioni giuridiche.

Un grave problema attinente al metodo classificatorio è che molti concetti giuridici sono vaghi e hanno un alone di connotazioni. Di conseguenza può essere molto difficile estrarre da tali concetti le condizioni necessarie e sufficienti per la loro applicazione, e distinguere da tali condizioni le caratteristiche accessorie la cui mancanza non impedisce l'applicazione del concetto. Così, di solito possiamo applicare con sicurezza un termine, solo quando consideriamo entità ben conosciute, che rappresentano istanze prototipiche di categorie preesistenti. E molto più rischioso affidarsi a classificazioni quando abbiamo a che fare con nuove entità, che per certi aspetti sono simili ai prototipi conosciuti, e per altri aspetti non lo sono. Tale rischio si aggrava quando si utilizzino categorie giuridiche molto ampie (soggetti, oggetti, persone, strumenti, ecc.), suscettibili di essere intese in accezioni diverse. Si consideri ad esempio come si possa affrontare in modi diversi il problema di stabilire se un'entità sia un soggetto o un oggetto per il diritto. Si può affermare che un'entità sia un soggetto se vale, alternativamente, anche una sola delle seguenti condizioni:

1. L'entità ha capacità giuridica: essa acquisterà diritti e doveri, qualora siano realizzate fattispecie atte a produrre quei diritti e doveri.

2. L'entità ha la capacità di compiere atti giuridici per se stessa: mediante le proprie dichiarazioni, essa può creare diritti e doveri per se stessa che corrispondono al contenuto di tali dichiarazioni.

3. L'entità ha la capacità di agire con effetti nella sfera giuridica altrui. Il diritto stabilisce che essa possa creare, mediante le proprie dichiarazioni, diritti e doveri in capo ad altri soggetti, anche se questi non avevano l'intenzione di effettuare tali dichiarazioni.

4. L'entità ha natura cognitiva o intenzionale. Vi sono conseguenze giuridiche (possibilmente in capo ad altri) che dipendono dal fatto che essa abbia certi stati cognitivi, conativi (scopi ed intenzioni) ed epistemici (informazioni o credenze).

5. L'entità è considerata come un centro di attività autonomo. La legge collega certe conseguenze al suo comportamento (che dipende dalla struttura interna dell'entità, e dal suo ambiente), anche quando questo comportamento non sia causato dall'azione o dall'omissione di una persona.

Come emerge da questo elenco, un'entità può essere un soggetto di diritto secondo una di tali descrizioni, senza essere un soggetto secondo altre descrizioni. Vi sono alcune connessioni logiche tra le nozioni di soggettività che abbiamo indicato, ma non si tratta di equivalenze. Chi parla di soggettività giuridica (ad esempio, al fine di attribuirle o di negarla agli AS o ad altri sistemi informatici) senza avere una mappa concettuale chiara e priva di pregiudizi, tenderà a commettere errori concettuali, passando inavvertitamente dall'uno all'altro concetto.

Ad esempio, nelle ipotesi di responsabilità oggettiva per i danni causati da una cosa, si può dire che la cosa sia un autonomo centro d'azione, cioè un soggetto nel senso sub (5). Tale entità non è un mero strumento (un'entità il cui comportamento rileva solo quando connesso con l'uso da parte di una persona), ma ciò non è sufficiente a farne un soggetto di tipo (4). A questo fine, come vedremo nel seguito, è necessario che l'entità possa possedere, o che possano esserle attribuiti, stati cognitivi o informativi. La soggettività di tipo (4) è connessa alla soggettività di tipo (3), nei limiti in cui chiamare un'azione una « dichiarazione », presuppone una competenza cognitiva da parte del dichiarante. La soggettività di tipo (3) non richiede né la soggettività di tipo (2) né quella di tipo (1): affinché le dichiarazioni (i negozi) compiute da un'entità creino diritti e doveri in capo a qualcun altro, non è necessario né che l'entità possa essere diretta titolare di quei diritti e doveri, né, a maggior ragione, che abbia la capacità di crearli da sé in capo a se stessa. Infine, la capacità di tipo (2) (la capacità di compiere atti giuridici per se stessi) presuppone la capacità di tipo (1) (la capacità di essere titolare di diritti o doveri) mentre non vale l'opposto (la capacità giuridica non necessariamente si accompagna alla capacità d'agire).

Le considerazioni appena riportate possono forse bastare a mostrare i pericoli di classificazioni affrettate, e condurci a concludere che, per fornire un adeguato modello giuridico per gli AS, assegnare etichette e assumere finzioni non è il metodo più appropriato. Al contrario, è necessario partire da un attenta analisi della natura della cosa di cui ci stiamo occupando, cioè delle proprietà e delle funzionalità che caratterizzano gli AS, e quindi individuare i tratti fondamentali di soluzioni giuridiche possano essere appropriate a tale natura. Solo dopo aver fatto ciò potremo esaminare per quali scopi e entro quali limiti possiamo qualificare gli AS secondo le categorie giuridiche tradizionali, e applicare loro la disciplina che è collegata a tali categorie.

2. UNA CARATTERIZZAZIONE DEGLI AS.

Non diremo molto circa la nozione generale di AS poiché si tratta di concetto discusso estesamente nella letteratura tec-

nica²? Ai nostri fini, può essere sufficiente affermare che un AS è un'entità computazionale che agisce autonomamente in un ambiente. Pertanto, ogni AS è inserito in un ambiente, percepisce l'ambiente (attraverso i suoi organi sensori), e può modificarlo (mediante i suoi organi effettori). L'attività dell'AS è autonoma in quanto determinata da processi cognitivi compiuti dallo stesso AS. Gli agenti elettronici possono operare sia nel mondo reale sia in quello virtuale, ma qui noi considereremo solo il secondo tipo di agenti (gli AS e non i robot fisici), e pertanto le percezioni e azioni degli agenti riguarderanno esclusivamente dati e messaggi informatici.

Le caratteristiche tipiche degli AS (benché esse possano presentarsi in diversa misura e in diverse combinazioni) sono le seguenti: l'inserimento (*embeddedness*) nel loro ambiente, la reattività e la pro-attività, la flessibilità comportamentale, la persistenza nel tempo, la mobilità, l'intelligenza, la capacità di comunicare. Tali caratteristiche comportano la difficoltà o anzi l'impossibilità di prevedere il comportamento degli AS. Questa impossibilità ha due aspetti distinti.

Il primo aspetto è quanto possiamo chiamare *imprevedibilità teorica*: la complessità degli AS (dovuta alla necessità di fornire le capacità che ricordavamo sopra: reattività, pro-attività, flessibilità, intelligenza, comunicazione, ecc.) combinata con la complessità dell'ambiente (o degli ambienti, per gli agenti mobili) rende una previsione accurata del comportamento degli AS molto difficile, e anzi impossibile in molte situazioni.

Il secondo aspetto è ciò che possiamo chiamare *imprevedibilità pratica*: per l'utilizzatore di un AS, dedicare le proprie energie all'esatta previsione del comportamento del proprio AS sarebbe in contraddizione con la ragione per la quale sta utilizzando l'AS, cioè delegare all'AS i compiti cognitivi implicati nell'attività affidata allo stesso AS³.

Consideriamo dapprima l'imprevedibilità teorica. Si consideri che l'utilizzatore di un AS non conosce i dettagli del funzionamento dell'AS, né conosce ogni dettaglio dell'ambiente nel quale l'AS opererà, né possiede le informazioni che l'AS ricercherà ed elaborerà al fine di svolgere i propri compiti. Pertanto, l'utilizzatore normalmente non sarà in grado di anticipare le azioni del suo AS. L'imprevedibilità esclude l'utilizzatore conoscesse o dovesse conoscere ogni azione compiuta dal suo AS. Essa esclude che ogni determinazione dell'AS corrisponda ad una determinazione del suo utilizzatore. Non possiamo dire: ogni qualvolta l'AS abbia

² Per un'introduzione generale alla nozione di AS, cf. Russeli and Norvig (1995), Weiss (1999), Wooldridge (2000).

³ Non cercheremo qui di precisare l'idea di delegazione, sulla quale, cf. Castelfranchi and Falcone (1998).

fatto alcunché, l'utilizzatore sapeva che l'avrebbe fatto, poteva impedire all'AS di farlo, e quindi voleva fosse fatto. Al massimo possiamo dire che in alcuni casi l'utilizzatore poteva aver previsto che l'AS avrebbe reagito in certi modi a certe possibili circostanze, di modo che egli aveva accettato che tali reazioni si sarebbero verificate (nell'eventualità che avessero luogo le circostanze corrispondenti).

Tuttavia anche una previsione condizionata è spesso impossibile. Ciò si verifica quando la vera ragione per la quale l'utilizzatore delega un compito ad un SA risiede nel fatto che l'utilizzatore non ha le abilità e le informazioni richieste per svolgere il compito affidato all'AS (ad esempio, l'utilizzatore di un *information agent* può non sapere dove cercare le informazioni di cui ha bisogno, o l'utilizzatore di un *negotiation agent* può non sapere chi contattare e come procedere nelle trattative in un particolare mercato). Si consideri altresì il caso di un AS che, oltre ad eseguire un certo compito, ad esempio, la compravendita di titoli azionari, sia capace di apprendere e di migliorare le proprie prestazioni. In tal caso, non solo l'utilizzatore non è in grado di prevedere che cosa farà l'AS (poiché egli non conosce le condizioni del mercato azionario in cui l'AS vi opererà), ma l'utilizzatore non è neppure in grado di prevedere le strategie che l'AS avrà appreso e quindi applicherà nel deliberare le proprie azioni. Quali strategie l'AS adotterà dipende da informazioni che solo l'AS avrà acquisito ed elaborato (informazioni attinenti all'evoluzione del mercato azionario), e attinenti ai risultati che lo stesso AS ha ottenuto nel passato utilizzando diverse strategie.

Passiamo ora a considerare l'imprevedibilità pratica, cioè il fatto che l'utilizzatore di un AS normalmente non ne anticipa il comportamento, né potrebbe essere ragionevolmente richiesto di farlo, anche quando l'anticipazione sia in teoria possibile. L'utilizzatore infatti impiega l'AS quale strumento cognitivo, strumento al quale egli ha delegato funzioni cognitive che ha scelto di non compiere direttamente: può trattarsi di funzioni epistemiche (accertare la presenza di certe condizioni o trovare certe informazioni) o di funzioni pratiche (scegliere che cosa fare sulla base delle informazioni disponibili, ad esempio, decidere se concludere un contratto e a quali condizioni). Sarebbe irragionevole richiedere che l'utilizzatore esegua direttamente, al fine di prevedere il comportamento del suo AS, proprio i compiti cognitivi che ha scelto di delegare al proprio AS: se così fosse la delegazione cognitiva sarebbe assurda o suicida (self-defeating). Ciò che l'utilizzatore può ragionevolmente cercare di prevedere è solo che il suo AS eserciterà correttamente queste funzioni cognitive, o almeno che le eserciterà abbastanza bene rispetto al compito ad esso affidato.

Gli AS sono soprattutto una tecnologia per delegare e decentralizzare i processi epistemici e deliberativi, e i loro utilizzatori vogliono liberare se stessi di questi compiti cognitivi (banali, noiosi,

costosi, ecc.). Pertanto, non ci dobbiamo sorprendere se gli AS eserciteranno tali funzioni non solo rispetto ad input cognitivi inaccessibili all'utilizzatore, ma anche e soprattutto rispetto ad input cognitivi che l'utilizzatore ha scelto di non elaborare direttamente (personalmente).

In conclusione, la finzione che l'utilizzatore conoscesse (o potesse conoscere) e pertanto volesse (o dovesse volere) quanto l'AS ha fatto o voluto, è contraddetta dalle due imprevedibilità appena considerate. Non solo è molto difficile ottenere tale conoscenza, ma investire tempo ed energie nell'ottenerla sarebbe irragionevole, dato che il vero scopo di chi utilizza l'AS è delegare funzioni cognitive. Non è richiamandosi a tale finzione che possiamo sperare di capire come il diritto disciplini e debba disciplinare l'uso degli AS.

Ricordiamo che la finzione dell'utilizzatore (o del programmatore) onnisciente è stata autorevolmente contestata, in relazione al più ampio tema dei contratti informatici, da Ettore Giannantonio (1997, 213 ss.)⁴. In particolare, questo autore, sulla base della sua profonda e diretta conoscenza delle tecnologie informatiche, rilevava l'insostenibilità della tesi secondo la quale la volontà nei contratti stipulati da sistemi informatici potrebbe essere imputata sempre all'uomo che ha preordinato i programmi che governano quei sistemi (come affermava Clarizia 1985). Giannantonio (1997, 250) osservava al riguardo che:

« se è vero che la capacità decisionale dell'elaboratore dipende dal programma, è anche vero che la gran parte delle decisioni che dal programma derivano non soltanto non sono conosciute da parte del programmatore, ma possono essere addirittura non previste e non prevedibili; e qualche volta possono essere perfino contrarie alle previsioni e alla volontà del programmatore ».

L'impossibilità di ricondurre il contenuto dei contratti informatici alla volontà dell'utilizzatore dell'elaboratore (o dei suoi ausiliari) portavano Giannantonio ad escludere che la volontà sia elemento o presupposto della contrattazione automatica poiché « nel caso di un regolamento di interessi concluso mediante l'elaboratore manca una vera e propria volontà nel senso voluto dal codice, ossia una volontà psicologica, cosciente, personale e attuale ».

Infine, queste considerazioni conducevano Giannantonio a configurare i contratti informatici quale nuovo modo di svolgimento dell'autonomia privata, consistente « non già in un comportamento, ma in un procedimento elettronico » (Giannantonio 1997, 257). Pur non essendo atti di volontà i contratti informatici

⁴ Il rigetto di questa finzione risponde all'esigenza di verità, rigore ed onestà intellettuale che aveva condotto Giannanto-

nio a contestare la presunzione di conoscenza della legge da parte del cittadino.

« sono considerati nella realtà sociale e degli affari vincolanti tra le parti e il giudice deve pertanto riconoscere la loro rilevanza giuridica e la loro forza obbligatoria ». Pertanto, nella prospettiva di Giannantonio, essi « costituiscono fonti di obbligazioni ai sensi dell'articolo 1173 del codice civile e sono soggetti alla disciplina delle obbligazioni contenuta nel titolo prima del libro quarto del codice civile ». Invece, ad essi non si applicherebbe la disciplina dei contratti (titolo secondo del libro quarto del codice civile), che riguarda le sole figure in cui « ricorra una dichiarazione di volontà e si sostanziano nella ricerca di un delicato equilibrio tra la realtà della dichiarazione e la realtà della volontà ».

Noi, pur condividendo pienamente la premessa dell'argomentazione di Giannantonio (l'impossibilità di ricondurre il contenuto dei contratti informatici ad una volontà umana), ne trarremo conclusioni parzialmente diverse: non solo i contratti informatici, quando conclusi da AS, costituiscono fonti di obbligazioni, ma essi sono disciplinati anche dalle norme sui contratti (seppur per via analogica). In particolare secondo la prospettiva che svilupperemo nelle pagine seguenti, le norme che fanno riferimento a stati psicologici (quali volontà, conoscenza, intenzione) possono essere applicate anche agli stati cognitivi degli AS. Giungeremo a tale conclusione partendo dall'osservazione che gli AS sono strumenti cognitivi, cioè strumenti ai quali gli utilizzatori delegano funzioni cognitive, e cui le controparti stesse riconoscono tali funzioni. Affermeremo infatti che gli effetti di quanto sia stato deliberato da un AS ricadono sull'utilizzatore dell'AS poiché l'utilizzatore ha scelto di utilizzare uno strumento cognitivo, e quindi si è assoggettato ai risultati dell'attività cognitiva di tale strumento.

Pertanto, nella nostra prospettiva, l'efficacia giuridica dell'azione dell'AS (specialmente in ambito contrattuale) dipende dalla volontà dell'utilizzatore (come ragionevolmente ricostruita dalle controparti dell'AS), ma si tratta della volontà di delegare certi compiti cognitivi all'AS, non della volontà di compiere tutte le specifiche azioni che saranno compiute dall'AS. Poiché l'utilizzatore intende impiegare le facoltà cognitive dell'AS, e ciò è noto alle potenziali controparti, il fatto che l'utilizzatore sia responsabile (nel senso che egli acquisterà gli obblighi e i diritti che risultano dall'attività dell'AS) non esclude, ma anzi presuppone la rilevanza giuridica degli stati e dei processi cognitivi dell'AS. Il fatto che gli AS abbiano propri stati cognitivi ed effettivo processi cognitivi non attribuibili all'utilizzatore (e il fatto che di ciò siano consapevoli tanto l'utilizzatore dell'AS quanto le controparti) distingue gli AS da altri oggetti o strumenti, anche dalla prospettiva del diritto.

È vero che il fenomeno che stiamo esaminando (la delegazione cognitiva ad un artefatto) non è completamente nuovo: tale delegazione si verifica in modo banale ogni qualvolta una persona impieghi una calcolatrice o un computer prima di concludere un contatto o adottare altra decisione giuridicamente rilevante (si consi-

deri il caso del negoziante che usi un computer per registrare vendite, calcolare prezzi e imposte, controllare carte di credito, ecc.). Tuttavia, di regola ciò riguarda solo i passi preliminari della deliberazione, e conduce a risultati cognitivi di cui si approprierà la persona che adotterà la deliberazione finale (la conclusione del contratto). Pertanto, di regola non c'è bisogno che il diritto consideri separatamente la cognizione automatica: quando, ad esempio, un software difettoso effettua un calcolo errato, del cui risultato l'utilizzatore si appropria, è sufficiente che il diritto prenda in esame l'errore dell'utilizzatore. Ciò non si verifica tuttavia, quando si affidi ad un AS il compito di svolgere direttamente un'attività giuridica, senza supervisione di persone fisiche, cioè quando, come dice l'art 12 di un recente Documento Uncitral⁵ «nessuna persona fisica abbia controllato ciascuna delle azioni eseguite da tali sistemi e l'accordo risultante»⁶. In tali circostanze, dobbiamo considerare direttamente le funzioni cognitive effettuate da sistemi informatici, e valutare quale debba essere la loro rilevanza giuridica, secondo la natura di tali sistemi, e le (ragionevoli) aspettative dei loro utilizzatori e interlocutori.

Si potrebbe affermare che anche la questione della cognizione artificiale non è completamente nuova per il diritto: c'è stato un progressivo sviluppo di macchine usate per compiere funzioni cognitive finalizzate a risultati giuridici, anche senza la supervisione di un utilizzatore: dai distributori automatici⁷, agli sportelli bancomat, alla contrattazione EDI, alla contrattazione mediante siti web. Questa constatazione, tuttavia, dovrebbe valere non come monito ad applicare anche agli AS le soluzioni adottate per le «macchine cognitive» già in uso, ma come invito a ripensare la disciplina giuridica di tutti gli strumenti cognitivi, e a individuare uno schema concettuale che possa applicarsi non solo agli strumenti più complessi, come gli AS, ma anche a strumenti più rudimentali, come quelli appena menzionati.

3. GLI STATI COGNITIVI NELLE ENTITÀ ARTIFICIALI.

Concentriamoci ora sul concetto di cognizione e consideriamo in quale misura questa nozione possa essere applicata agli AS e ad altri artefatti, da una prospettiva giuridica.

⁵ Legal aspects of electronic commerce. Electronic contracting: provisions for a draft convention. United Nations Commission on International Trade Law. Working Group iv (Electronic Commerce). Thirty-ninth session. New York, 11-15 marzo 2002).

⁶ «No natural person reviewed each of the individual actions carried out by such systems or the resulting agreement».

⁷ Rispetto ai quali, come è noto, il primo riferimento nella dottrina italiana è Cicu (1901).

Il nostro punto di partenza sarà rappresentato dalla classica distinzione proposta da Daniel Dennett (1989; 1991; 1997) tra la prospettiva (*physical stance*) fisica, la prospettiva progettuale (*design stance*) e la prospettiva intenzionale (*intentional stance*)⁸. Le tre prospettive individuano tre modi nei quali si può guardare al comportamento di entità naturali o artificiali:

1. come determinato dalle leggi fisiche applicabili (la prospettiva fisica),
2. come determinato dalle funzioni che tali entità e le loro componenti sono destinate a svolgere (la prospettiva progettuale),
3. come determinato dagli stati cognitivi (intenzionali) che tali entità posseggono (la prospettiva intenzionale).

Il fatto di guardare ad un'entità da una prospettiva non esclude che si possano adottare anche le altre prospettive. Pertanto si possono attribuire ad un essere umano credenze e desideri (prospettiva intenzionale), ma si possono anche considerare le funzioni che diverse aree del cervello e parti del corpo svolgono nella cognizione (prospettiva progettuale), e si può infine studiare la chimica delle interazioni sinaptiche nel sistema nervoso (prospettiva fisica). Allo stesso modo possiamo assumere che un sistema automatico stia perseguendo certi scopi, sulla base delle informazioni che esso possiede (prospettiva intenzionale), ma possiamo anche considerare le funzioni svolte dalle diverse componenti software del sistema e i modi in cui queste componenti interagiscono (prospettiva progettuale), e infine possiamo studiare le dinamiche elettriche dell'hardware dello stesso sistema (prospettiva fisica).

Tutte e tre le prospettive sono non solo legittime, ma anche utili, nelle condizioni opportune. Dobbiamo spostarci verso l'alto nella scala dell'astrazione (dalla prospettiva fisica, a quella progettuale, a quella intenzionale) per accostarci ad entità complesse pur senza analizzarne compiutamente il funzionamento interno, cioè per comprenderne il comportamento in termini generali, anche quando non abbiamo una conoscenza completa della struttura interna di tali entità e delle loro interazioni causali con l'ambiente. Dobbiamo scendere nella scala dell'astrazione per individuare il meccanismo che conduce il sistema ad avere un certo macrocomportamento, e in particolare, per individuare le cause del suo cattivo funzionamento (dobbiamo aprire il cofano dell'automobile per spiegare il suo improvviso arresto, dobbiamo analizzare il co-

⁸ Qui il termine « intenzionale » è utilizzato nel significato tecnico che esso ha nel linguaggio filosofico, dove tipicamente si riferisce alla relazione (chiamata anche « aboutness ») tra gli stati mentali o gli oggetti linguistici e le cose cui si riferiscono.

Non stiamo usando il termine « intenzionale », nel significato familiare di « deliberato » o « volontario ». Pertanto anche credenze, desideri, speranze e paure sono stati intenzionali. Sull'intenzionalità, cf. Dennett and Haugeland (1987).

dice del programma per spiegare l'improvviso blocco della sua esecuzione, ecc.).

Concentriamoci ora sulla prospettiva intenzionale⁹. Quando osserviamo un'entità dalla prospettiva intenzionale, spieghiamo il comportamento di quell'entità assumendo che essa abbia certi stati cognitivi. Tra gli stati cognitivi includiamo tanto gli stati epistemici (il possesso di informazioni sull'ambiente) quanto gli stati conativi (il possesso di determinazioni che guidano l'azione). Tipicamente, assumiamo che l'entità che stiamo esaminando stia cercando di raggiungere certi obiettivi (goal) o di applicare certe istruzioni (intenzioni) sulla base di certe rappresentazioni del suo ambiente (credenze). Il comportamento di un'entità intenzionale sarà poi spiegato in quanto corrispondente all'ipotesi che l'entità realizzi le proprie intenzioni o persegua i propri scopi (mediante mezzi che essa ritiene appropriati).

Ad esempio, per capire e prevedere il comportamento di un sistema informatico per il gioco degli scacchi di regola non si può adottare né la prospettiva fisica (determinare l'evoluzione degli stati elettrici delle componenti del computer su cui il sistema è installato), né la prospettiva progettuale (considerare le funzioni rivestite dai moduli del software e dalle istruzioni di programmazione che compongono quei moduli). Normalmente l'avversario di un sistema siffatto non sa nulla della sua struttura software e ancor meno dello stato elettronico dei suoi componenti hardware. Inoltre, anche un programmatore che abbia collaborato allo sviluppo del sistema non può anticiparne il funzionamento eseguendo mentalmente le istruzioni di programmazione che esso comprende (il sistema è troppo complesso perché ciò risulti fattibile). Possiamo prevedere il comportamento del sistema solo attribuendo ad esso determinati obiettivi (il sistema « vuole » vincere la partita, attaccare certi pezzi, raggiungere una certa posizione), e determinate conoscenze (il sistema « sa » che certe mosse e strategie sono disponibili, che certe mosse o strategie possono essere adottate dall'avversario), e assumendo che esso possa individuare metodi appropriati (razionali) per raggiungere i propri obiettivi alla luce delle proprie conoscenze¹⁰.

Come illustra l'esempio appena presentato, la prospettiva intenzionale è il solo modo in cui possiamo sperare di comprendere e

⁹ Per una discussione delle implicazioni giuridiche dell'adozione della prospettiva intenzionale, cf. Sartor (2002b).

¹⁰ Secondo Dennett (1989, 17), la prospettiva intenzionale opera come segue: « first you decide to treat the object whose behaviour is to be predicted as a

rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs ».

prevedere l'azione di un sistema informatico complesso, come in particolare un AS. Tale previsione non può essere fondata sull'analisi del meccanismo computazionale che costituisce l'AS, e sulla predeterminazione delle relazioni di questo meccanismo a tutti gli input possibili. L'utilizzatore di un AS normalmente non sarà a conoscenza di questo meccanismo, e anche il programmatore che abbia partecipato allo sviluppo dell'AS sarà incapace di riconoscere nel comportamento dello stesso l'esecuzione dei precisi processi computazionale. L'interpretazione complessiva del comportamento dell'AS dovrà invece basarsi sulla prospettiva intenzionale, cioè sull'ipotesi che l'AS stia operando razionalmente, adottando determinazioni appropriate rispetto agli scopi ad esso assegnati, nel contesto in cui si trova ad operare. Quest'assunzione di razionalità non è assoluta. Al contrario, essa può essere integrata dalla conoscenza delle limitazioni delle capacità dell'AS (in modo che sia possibile spiegare perché l'AS non riesca a comportarsi razionalmente in circostanze particolari). Questo modello esplicativo è simile alla strategia che adottiamo rispetto agli esseri umani: possiamo interpretare e prevedere il comportamento di altre persone combinando l'ipotesi generale della loro razionalità con la conoscenza delle limitazioni e delle idiosincrasie di ogni individuo.

Se la prospettiva intenzionale deve essere adottata dall'utilizzatore di un AS, a maggior ragione essa deve essere adottata dalla controparte dell'AS negli scambi in cui l'AS agisce per conto del suo utilizzatore/proprietario. La controparte dell'AS non può neppure provare a comprendere il comportamento dell'AS analizzando le componenti software (il codice del software essendo di regola inaccessibile, e in ogni caso troppo complesso per essere studiato in tempo utile), né interrogandosi su quali intenzioni dell'utilizzatore, codificate nel software, l'AS possa esprimere.

Si consideri ad esempio, un commesso virtuale animato, che appare come un'immagine tridimensionale dotata di *body language* (espressioni facciali, gesti, ecc.) e parola, e introduce il cliente in un negozio virtuale (di antiquariato, automobili usate, ecc.), presentandogli i prodotti, interrogandolo sui suoi bisogni, suggerendogli certe scelte, e proponendogli certe clausole contrattuali. Si consideri anche l'ipotesi di un tour-operator virtuale, che parla attraverso il telefonino dell'utilizzatore, offrendogli di acquistare certi biglietti, a certe condizioni. Infine, si consideri un AS che opera in un mercato dinamico, e che contatta persone e altri AS al fine di trovare le opportunità di scambio più convenienti. Per gli interlocutori di tali AS, la sola chiave per capirne il comportamento sarà l'ipotesi che gli AS, al fine di raggiungere obiettivi loro assegnati, e usando la conoscenza di cui dispongono, raggiungano determinazioni che dichiarano alle controparti, secondo le convenzioni linguistiche e sociali esistenti. Come abbiamo appena osservato, l'assunto della razionalità degli AS deve essere integrato con la conoscenza dei limiti di tale razionalità, ma tale assunto for-

nisce comunque il presupposto per capire il comportamento degli stessi AS.

L'idea della prospettiva intenzionale, come risulta dal modello proposto da Dennett sembra condurci verso la conclusione che gli stati cognitivi esistano solo nell'occhio dell'osservatore. Essi sembrano consistere in un modo particolare di guardare ad un'entità, che non è traducibile in caratteristiche interne a tale entità. Affermare che un'entità ha certi stati cognitivi (obiettivi, informazioni, credenze, desideri, ecc.) consisterebbe nell'affermare che il comportamento di tale entità risulta spiegabile e prevedibile secondo la prospettiva intenzionale, cioè mediante l'attribuzione di stati cognitivi (e postulando che l'entità possa comportarsi razionalmente sulla base di tali stati cognitivi). Così, perché sia vero che il mio scacchista automatico « vuole » mangiarmi la torre sembra sufficiente che, attribuendo questo obiettivo allo scacchista (e assumendo che esso sia in grado di agire in modo da raggiungere i suoi obiettivi) io possa prevederne il comportamento successivo (anticiparne le future mosse).

Tuttavia, è anche possibile una diversa impostazione. Si può adottare l'impostazione realistica secondo la quale gli asserti relativi a stati cognitivi si riferirebbero a specifiche caratteristiche interne dell'entità cui sono attribuite. Così, per stabilire se un'entità possieda realmente stati cognitivi (scopi, credenze, intenzioni, ecc.) bisognerebbe considerare se vi siano condizioni interne a tale entità che rappresentano stati epistemici e conativi, e se vi siano modi di funzionamento di tale entità che realizzano tecniche razionali per elaborare informazioni epistemiche e conative. Il comportamento esterno di tale entità quindi rilevarebbe (ai fini del possesso di stati intenzionali) solo quale indizio del suo funzionamento interno.

Seguendo questa linea di pensiero, possiamo affermare che lo stato interno di una certa entità (ad esempio, la presenza di una certa sostanza chimica nel sistema circolatorio di un organismo, o la presenza di certe sequenze di caratteri in una certa variabile, o in un certo *buffer* di dati di un sistema informatico) rappresenta uno stato epistemico, e più precisamente, la credenza nell'esistenza di certe situazioni, quando:

- l'entità interessata tenda ad adottare tale stato in presenza di tali situazioni (in tali situazioni i sensori dell'entità vengono attivati, e ciò dà inizio ad un processo causale che conduce l'entità ad adottare lo stato in questione), e inoltre

- il fatto che l'entità interessata si trovi in tale stato contribuisce a fare in modo che l'entità stessa si comporti come le situazioni corrispondenti richiedono.

Pertanto, lo stato interno di un'entità è la credenza nell'esistenza di certe situazioni esterne (o, se si preferisce, rappresenta o indica tali situazioni all'entità), quando:

- vi è una covarianza tra lo stato interno e tali situazioni, e

• tale covarianza ha la funzione di fare in modo che l'entità possa reagire appropriatamente alla presenza di tali situazioni.

Non possiamo approfondire qui il difficile problema della natura dell'intenzionalità¹¹. Limitiamoci ad osservare che possiamo guardare da questa prospettiva a sistemi informatici e in particolare ad AS. Attribuire stati epistemici a sistemi informatici consentirebbe, almeno in alcuni casi, di trovare una disciplina giuridica appropriata senza modifiche legislative, e inoltre consentirebbe di distinguere chiaramente le ipotesi in cui il fatto che un sistema possieda, o faccia in modo che altri posseggano, certi stati epistemici, sia rilevante per il diritto.

Si consideri ad esempio l'azione di inserire un nome nella casella riservata al nome del compratore, al fine di ordinare un prodotto da un sito web. La registrazione del dato inserito in tale casella certamente tende a covariare con nome della persona che sta facendo l'ordine d'acquisto, e ha la funzione di far sì che il sito si comporti in un modo appropriato nei confronti di un vero acquirente (e non di una persona che usa l'identità altrui senza autorizzazione). Pertanto, possiamo dire che il sito « crede » che il nome registrato sia il nome della persona che ha effettuato la registrazione (o ha autorizzato l'uso del proprio nome a tal fine). Quando il nome indicato sia diverso dal nome reale, possiamo dire che il sito è stato ingannato, cioè che esso è stato indotto ad avere una falsa rappresentazione della realtà. Ciò consentirebbe di applicare, almeno analogicamente, le regole giuridiche relative al dolo e alla frode, anche alle interazioni con sistemi informatici (un passo che, invece, la maggior parte dei giuristi sono stati restii a fare, il che ha richiesto una disciplina specifica, spesso assai rozza formulata, sulla frode informatica, come quella di cui all'art. 604-ter c.p.).

L'idea che un computer (un software) possa avere stati cognitivi può essere estesa anche agli stati conativi. Si può affermare che un'entità ha lo scopo di realizzare un certo risultato (in termini più antropomorfici, che essa ha il desiderio di realizzarlo) quando vi sia uno stato interno dell'entità tale che:

- finché l'entità continua a trovarsi in tale stato, essa tende a realizzare il risultato, e
- quando il risultato è raggiunto, lo stato interno viene abbandonato (o modificato in modo che esso cessi di determinare il comportamento di cui sopra).

Ad esempio, si assuma che si sia fornita ad un AS la descrizione di certi beni, e che, sulla base di tale descrizione, l'AS inizi ad at-

¹¹ « Sull'idea di covarianza, cf. Dretske (1986), e per una discussione della letteratura, cf. Davies (1998, 287) ss. »

tivarsi per acquistare quei beni, e continui finché i beni non siano stati acquistati (il che determinerà la rimozione della descrizione dei beni dalla lista dei compiti che l'AS deve eseguire). Date queste condizioni, possiamo certamente affermare che l'AS ha l'obiettivo, scopo, o fine, di acquistare quei beni.

Possiamo anche dire che un sistema voleva compiere una certa azione, se c'era un'istruzione nel sistema che prescriveva al sistema stesso di compiere quell'azione. Inoltre, possiamo dire che il sistema « voleva » adottare un certo comportamento, se il sistema adottò quel comportamento sulla base di un processo interno, inteso a fare in modo che il sistema raggiungesse i propri scopi, sulla base dei propri stati epistemicici.

Ad esempio, se un AS ha lo scopo di causare un risultato dannoso (ad esempio, l'interruzione del funzionamento di un sistema informatico) e sceglie di compiere un'azione che produce quel risultato, quale modo di produrre quel risultato, non solo possiamo dire che l'AS vuole compiere quell'azione, ma possiamo anche affermare che esso vuole produrre il risultato, cioè che l'AS causa deliberatamente il danno.

Concludiamo questa breve discussione degli stati cognitivi nei sistemi informatici affermando che, dal punto di vista del giurista, non c'è necessità di effettuare una scelta tra le due prospettive appena esaminate, cioè tra il considerare gli stati cognitivi:

- come interpretazioni del comportamento esteriore di un sistema, e
- come condizioni interne al sistema stesso.

Le due prospettive sono, innanzitutto, collegate da un nesso causale: di regola un'entità può operare in un modo che corrisponda ad una certa interpretazione cognitiva, proprio in quanto essa possieda stati interni del tipo appena descritto. Inoltre, da una prospettiva giuridica, le due concezioni sono complementari. Da un lato, l'idea degli stati cognitivi come interpretazioni del comportamento di un'entità fa riferimento alle attitudini degli osservatori esterni (quali credenze, scopi e intenzioni le attribuiscono le controparti e i terzi?). Dall'altro lato l'idea degli stati cognitivi quali stati interni fa riferimento al punto di vista dell'entità stessa, o di coloro che possono ispezionarne il funzionamento interno (credeva realmente l'entità quanto la controparte riteneva essa credesse, e aveva gli scopi che la controparte assumeva essa avesse?).

Pertanto, possiamo concludere affermando che, secondo il modello da noi proposto, l'attribuzione di stati cognitivi ad un'entità si fonda sull'ipotesi che l'entità possieda la capacità di agire razionalmente (efficientemente) per raggiungere i propri obiettivi e attuare le proprie intenzioni, sulla base delle informazioni che può estrarre dal proprio ambiente. Questa capacità si basa, tuttavia, sul fatto che l'entità in questione possieda appropriati stati interni (cognitivi).

Pertanto, gli asserti che attribuiscono ad un AS uno stato cognitivo sono veri quando l'AS possiede uno stato interno che svolge la funzione dello stato cognitivo attribuito all'AS. Ad esempio, l'asserto che un AS ha una credenza in una certa proposizione (ha certe informazioni) sarà vero quando vi sia uno stato interno dell'AS che rappresenta tale proposizione e che funziona appropriatamente nel processo deliberativo dell'AS. Gli stessi asserti sono ragionevoli o giustificati quando vi siano sufficienti indizi per concludere che l'AS possiede lo stato cognitivo che è ad esso attribuito. Ad esempio, quando un AS accetta un certo tipo di input cognitivo (ad esempio, l'informazione che un certo oggetto è fatto di un particolare materiale), e quindi si comporta in un modo appropriato, posso concludere che esso ha una credenza che corrisponde all'input ad esso fornito.

La nostra discussione del problema dell'intenzionalità dei sistemi informatici ci conduce alla conclusione che gli AS (come altri sistemi informatici) possono avere stati cognitivi e possono effettuare processi cognitivi. Considerarli in questo modo è, in molti casi, il solo modo appropriato di accostarci ad essi. Pertanto delegare loro compiti cognitivi è il modo appropriato di utilizzarli, e assumere che ad essi siano stati delegati tali compiti è il modo appropriato per interagire con essi. Nelle pagine seguenti considereremo le implicazioni giuridiche di ciò, rispetto a diverse aree del diritto. Non considereremo approfonditamente tutte le possibili questioni¹², ma ci limiteremo ad esaminare alcuni ambiti nei quali gli atteggiamenti cognitivi sono particolarmente rilevanti.

4. AS E RESPONSABILITÀ CIVILE.

Alcuni problemi attinenti alla responsabilità per i danni causati da un AS sono comuni ad altri oggetti tecnologici.

Consideriamo, innanzitutto, il problema di identificare un custode, al fine di renderlo responsabile per i danni causati da AS. Rispetto agli oggetti materiali tale identificazione è di regola facile: il custode è il proprietario, sempre che questi non abbia trasferito il controllo della cosa ad un terzo (ad esempio, il comodatario o il mutuatario). Tuttavia, circa gli AS, dobbiamo considerare che diversi soggetti possono « possedere » o « controllare » diversi aspetti di un AS. Se l'AS attua una tecnologia oggetto di brevetto, il controllo su questa non potrà che competere al titolare del brevetto. Se l'AS incorpora, come accade di regola, un software

¹² Per una rassegna dei principali problemi del diritto degli AS, cf. Cevenini et al (2002).

coperto da diritto d'autore, allora il controllo sul software, inaccessibile all'utilizzatore, spetterà al titolare del diritto d'autore. Se l'AS comprende una banca di dati, allora i dati saranno sotto il controllo di chi li abbia raccolti e organizzati. Infine, se l'AS opera seguendo le indicazioni di un utilizzatore, allora si può dire che l'utilizzatore controlla la particolare combinazione di tecnologia, software e dati, che costituisce l'AS¹³. In ogni caso, poiché l'utilizzatore non ha controllo su certi aspetti del suo AS, può essere iniquo ritenerlo responsabile in relazione ai danni collegati a questi aspetti. Ad esempio, l'utilizzatore dovrebbe essere un custode rispetto ai difetti del software, quando non ha accesso al codice sorgente, e il diritto stesso gli vieta di decodificare e modificare il sorgente. Per accostarsi al tema della responsabilità per i difetti di funzionamento collegati a tali aspetti, la stessa idea di custodia è inadeguata: o la estendiamo, in modo che ricopra anche la responsabilità di produttori, progettisti e sviluppatori (questa direzione, come è noto, è stata seguita nel diritto francese), o integriamo la responsabilità civile con il riferimento ad altri settori del diritto (responsabilità del produttore, tutela del consumatore, ecc.).

Un secondo problema, ancora comune ad altri oggetti tecnologici, concerne l'ambito della responsabilità del custode. Come è noto, sono possibili due costruzioni di questa forma di responsabilità. Secondo la prima, il custode è responsabile solo se, e nella misura in cui, ha negligenzemente omesso di controllare la cosa. Tuttavia, può essere molto difficile identificare una mancanza colpevole di controllo da parte dell'utilizzatore di un AS, poiché gli AS hanno la possibilità di agire al fuori del controllo dei loro utilizzatori, e in modi che questi non potevano prevedere. Se seguiamo questa prospettiva, allora dobbiamo concludere che in molti casi (e probabilmente nella maggior parte di essi), nessuno sarebbe responsabile per i danni causati da AS. Conseguentemente, tutti gli utilizzatori di Internet dovrebbero essere esposti al rischio di possibili danni, in seguito al comportamento imprevisto degli AS. Ciò potrebbe contribuire a compromettere la fiducia nelle rete e, data la difficoltà di provare la mancanza di controllo, potrebbe fornire uno scarso incentivo all'uso responsabile delle tecnologie basate sugli agenti.

Una prospettiva alternativa consisterebbe nell'assumere che il custode di un AS sia sempre responsabile per ogni danno causato da un AS, a prescindere dalla violazione di un obbligo di diligenza. Questo modello (argomentabile anche con riferimento all'art. 2050 c.c.) consentirebbe di trasferire le perdite economiche

¹³ Sulla proprietà e il controllo degli AS, dal punto di vista dell'informatica, cf. Pitt et al (2001), Yip e Cunningham (2002).

(causate dagli AS) dalla persona danneggiata al custode. Questa soluzione tuttavia potrebbe sembrare eccessivamente dura nei confronti del custode (l'utilizzatore) che affronterebbe una responsabilità imprevedibile, anche per eventi al di là del suo controllo. Tuttavia, qualche aiuto nell'allocazione della responsabilità può essere fornito dagli standard suggeriti dall'analisi economica del diritto: la responsabilità va posta in capo al soggetto che possa evitare i danni (o assicurarsi contro di essi) con minore spesa. Secondo questo criterio la responsabilità dovrebbe essere posta, a seconda del tipo di problema che ha causato il danno, a carico dello sviluppatore, del proprietario, o dell'utilizzatore dell'AS. Inoltre, bisogna prendere in considerazione la possibilità di far gravare il danno, almeno in parte, sul danneggiato che abbia contribuito al suo verificarsi (a norma dell'art. 1227 c.c.).

Sin qui siamo rimasti nell'ambito di temi giuridici ben noti, rispetto ai quali le questioni collegate agli AS non sono diverse da quelle che possono presentarsi rispetto ad altri oggetti tecnologici.

Le nuove questioni che dobbiamo affrontare attengono alla caratteristica degli AS che abbiamo introdotto sopra, il fatto che essi sono strumenti cognitivi. Dobbiamo considerare, cioè se gli stati cognitivi di un AS possano essere rilevanti per stabilire e circoscrivere le responsabilità derivanti un danno causato da quell'AS. Si osservi che porsi questo quesito non significa chiedersi se l'AS sia giuridicamente o moralmente responsabile. Ciò è anzi irrilevante ai nostri fini, poiché per « responsabile » (giuridicamente, moralmente, o sotto altri profili) qui intendiamo « obbligato a risarcire il danno », e la sola responsabilità che qui stiamo considerando è la responsabilità dell'utilizzatore. Tuttavia, il fatto che la responsabilità dell'utilizzatore possa dipendere dagli stati cognitivi del suo AS, differenzia gli AS da altre cose e strumenti, e comporta un'analogia con la responsabilità vicaria.

Questo aspetto emerge con chiarezza quando dobbiamo stabilire quali eventi siano stati « causati » da un AS. Ciò è importante in particolare se riteniamo che vi sia una responsabilità oggettiva a carico dell'utilizzatore dell'AS, poiché in questo caso il fatto che l'AS abbia causato il danno sarà fondamento sufficiente della responsabilità dell'utilizzatore.

Si consideri ad esempio il caso di un AS che invii un messaggio in apparenza innocente ad un sistema informatico (« prezzo offerto euro 75 »), e si assuma che questo messaggio inizi un processo che conduce al blocco del sistema destinatario, a causa di un difetto di quel sistema. Si assuma altresì che la ricezione di quel messaggio sia stata necessaria affinché si verificasse il blocco (senza il messaggio il blocco del sistema non avrebbe avuto luogo). Per allocare la responsabilità dobbiamo rispondere alla seguente domanda: fu il blocco causato dal messaggio dell'AS (di modo che l'utilizzatore dell'AS, quale suo custode, debba risarcire il danno), oppure fu il blocco causato da un difetto interno al si-

stema (il messaggio si limitò a fornire l'occasione affinché il difetto interno operasse)? Come è noto un criterio per limitare la causalità è quello della casualità « normale » o « adeguata » (non possiamo affrontare qui il difficile tema della causalità, uno dei concetti giuridici più controversi): un evento « causa » solo gli effetti che normalmente seguono da eventi di quel tipo, mentre l'effetto eccezionale di un evento, dovuto al concorso di fattori eccezionali (come il cattivo funzionamento del sistema interlocutore, nel nostro esempio) non si considera causato dall'evento.

La limitazione della causalità agli effetti « normali », tuttavia, può portare a risultati assurdi nel caso in cui i danni siano stati intenzionalmente (deliberatamente) prodotti, anche quando l'intenzione in questione sia l'intenzione di un AS, anziché quella di una persona.

Consideriamo due diversi casi ipotetici. Nel primo caso, l'AS che ha inviato il messaggio era a conoscenza dell'esistenza di un difetto nel sistema destinatario, e ha inviato il messaggio proprio al fine di causare il blocco di quel sistema. Nel secondo caso, l'AS ha inviato il messaggio in buona fede, al fine di fare una proposta di acquisto. I due casi ipotetici sono identici rispetto al comportamento degli AS implicati, che consiste in un comportamento normalmente innocente: inviare il messaggio « prezzo offerto euro 75 ». La sola differenza tra loro consiste nelle ragioni che hanno motivato gli AS ad inviare il messaggio.

Pertanto, vi sono due possibilità: o trattiamo i due casi ipotetici nello stesso modo, o dobbiamo distinguerli sulla base dei diversi stati cognitivi degli agenti in essi implicati (assumiamo che sia disponibile un modo per ricostruire tali stati cognitivi). Se i due casi debbono essere trattati nello stesso modo, dovremmo emettere un giudizio di non responsabilità (per gli utilizzatori degli agenti) in entrambi i casi. Ma ciò incentiverebbe la costruzione di AS che tendano a sfruttare i difetti di altri sistemi, poiché l'utilizzatore di tali agenti non sarebbe mai responsabile per comportamenti normalmente innocenti dei suoi agenti (anche quando quei comportamenti siano stati concretamente malevoli). Pertanto, sembra dobbiamo differenziare la disciplina giuridica dei due casi ipotetici: il danno intenzionalmente causato da un AS dovrebbe determinare la responsabilità del suo utilizzatore, anche se prodotto in circostanze eccezionali (specialmente quando tali circostanze sono note all'AS), mentre il danno causato non intenzionalmente non dovrebbe produrre tale risultato nelle stesse circostanze. Più in generale, seguendo l'impostazione da ultimo enunciata, un utilizzatore sarebbe responsabile per i danni prodotti « deliberatamente » dal suo AS (i danni che l'AS intendeva realizzare o che esso abbia previsto quali effetti della propria azione), e per i danni che l'AS abbia prodotto violando doveri di diligenza che riguardano l'attività che l'AS stava effettuando (danni che l'AS avrebbe dovuto prevedere ed evitare).

Si osservi che l'idea che un AS possa disattendere doveri di diligenza non implica che l'AS sia responsabile (nel senso di essere soggetto a sanzione) per la violazione di quei doveri. Tale idea implica solo che se l'AS non si conforma a tali doveri (esso non anticipa gli effetti probabili del proprio comportamento e non usa appropriate cautele) allora l'AS ha funzionato male (come strumento cognitivo) di modo che il suo utilizzatore sarà responsabile, come utilizzatore (o proprietario) di una macchina difettosa. Dall'altro lato, se l'AS ha usato tutta la cautela e la diligenza richiesta dall'attività da esso svolta, allora l'utilizzatore non dovrebbe essere responsabile del danno risultante dall'attività dell'AS, poiché in tal caso non vi è stato alcun mal-funzionamento (sempre che non ci si trovi di fronte ad una situazione nella quale anche alle azioni dello stesso utilizzatore si applicherebbe un regime di responsabilità oggettiva).

Pertanto, sembra che la responsabilità del custode per l'azione di un AS non possa essere fondata solo sul fatto che un danno poteva essere previsto secondo le « normali » leggi della natura (o della tecnica). Dobbiamo inoltre considerare anche se l'AS abbia intenzionalmente o negligenzemente prodotto il danno. Pertanto, la responsabilità dell'utilizzatore di un AS viene ad essere assimilata, anziché alla responsabilità del custode di una cosa, alla responsabilità vicaria (la responsabilità di padroni e committenti, *ex* 2049 c.c., e in particolare, del datore di lavoro per i fatti dei dipendenti). Come è noto questa forma di responsabilità civile non è fondata sul fatto che il datore di lavoro potesse prevedere il comportamento del dipendente, ma invece sul fatto che il dipendente abbia commesso un illecito nell'ambito della sua attività lavorativa.

Si osservi che la rilevanza degli stati cognitivi dell'AS è la ragione sola per la quale la relazione tra un AS e il suo utilizzatore può essere assimilata, in certe circostanze, a quella tra un dipendente e il suo datore di lavoro (ai fini della responsabilità vicaria). Ciò non presuppone che l'AS sia classificato come una persona, o come soggetto morale o giuridico.

5. AGENTI E CONTRATTI.

La contrattazione (e la trattativa) on-line è già oggi un'ambito di applicazione molto importante per le tecnologie basate sugli agenti (sulla disciplina giuridica dei contratti stipulati da agenti, cf. Lerouge 2000, Weitzenboeck 2001a, DeMiglio et al 2002). Ciò è confermato dal fatto che alcuni legislatori hanno già mostrato un certo interesse per questo tema. In particolare, l'Uniform Computer Information Transactions Act (UCITA) statunitense stabilisce alcune regole che riguardano specificamente gli agenti elettronici. L'UCITA definisce l'agente elettronico come « un programma per ela-

boratore o uno strumento elettronico o comunque automatico, usato per iniziare un'azione indipendentemente, o in risposta a messaggi o prestazioni elettroniche, per conto della persona, senza supervisione dell'azione da parte di un individuo al tempo dell'azione o della risposta al messaggio o prestazione »¹⁴ e (nella sezione 107 d)) afferma che « la persona che usa un agente elettronico che ha scelto per effettuare un'autenticazione, prestazione o accordo, inclusa la manifestazione del consenso, è vincolata dalle operazioni dell'agente elettronico, anche se nessun individuo era a conoscenza o verificava le operazioni dell'agente o i risultati delle operazioni »¹⁵. Simili indicazioni sono contenute nel documento Uncitral sopra menzionato, che suggerisce che una nuova convenzione sulla contrattazione elettronica dovrebbe prevedere espressamente la validità di contratti stipulati mediante agenti automatici, anche senza controllo umano, e afferma che l'utilizzatore è responsabile per le dichiarazioni dell'agente, anche quando non abbia voluto tali dichiarazioni.

Anche con riferimento ai contratti conclusi da agenti, ci occuperemo delle applicazioni più « normali », tralasciando ipotesi improbabili o fantascientifiche. Quindi non ci chiederemo neppure se AS operanti nel proprio interesse possano essere considerati quali soggetti autonomi, dotati di una soggettività morale o giuridica, e possono entrare in accordi giuridicamente vincolanti in nome proprio. Ci limiteremo agli AS incaricati di trattare e stipulare contratti nel nome e nell'interesse dei propri utilizzatori. Si tratta delle ipotesi nelle quali l'AS acquisti o venda beni per conto del proprio utilizzatore, e più in generale, delle ipotesi nelle quali esso effettui attività attinenti alla formazione di offerte contrattuali o alla loro accettazione. Sosterremo che per affrontare in modo coerente questi contesti dobbiamo accettare la tesi che l'AS, pur non essendo un soggetto di diritto, possa aver stati cognitivi rilevanti per il diritto.

Innanzitutto, come abbiamo osservato nella sezione 1, dobbiamo respingere l'opinione che gli AS si limitino a trasmettere contratti preparati dal loro utilizzatore (o programmatore). Questa opinione, come osservavamo sopra, è incompatibile con il fatto che né l'utilizzatore né il programmatore sono in condizione da anticipare completamente il comportamento contrattuale dell'AS

¹⁴ « A computer program or electronic or other automated means, used independently to initiate an action, or to respond to electronic messages or performances, on the person's behalf without review of action by an individual at the time of the action or response to the message or performance ».

¹⁵ « A person that uses an electronic agent that it has selected for making an authentication, performance or agreement, including manifestation of assent, is bound by the operations of the electronic agent, even if no individual was aware or reviewed the agent's operations or the results of the operations ».

in tutte le circostanze possibili, e pertanto non sono in grado di « volere » i contratti che saranno conclusi dall'AS. Abbiamo anche osservato che, anche quando l'utilizzatore sia in condizione di effettuare tale previsione, egli non può ritenersi obbligato (o onerato) ad effettuarla, poiché ciò sarebbe in contraddizione con la ragione per utilizzare un AS: delegare compiti cognitivi, come l'acquisizione della conoscenza e la deliberazione pratica (la scelta di fare un contratto con un certo contenuto). Pertanto, il fatto che gli effetti di un contratto stipulato da un AS ricadano sull'utilizzatore non può essere spiegato con il fatto che l'utilizzatore abbia previsto il comportamento del suo AS, o potesse prevederlo, o dovesse prevederlo (come affermato, per esempio, da Finocchiaro 2002). È vero che un'intenzione dell'utilizzatore (come riconoscibile alla controparte) fornisce la giustificazione ultima del valore giuridico dei contratti stipulati dal suo AS, ma si tratta dell'intenzione di affidare all'AS il compito di entrare in rapporti di scambio per conto dell'utilizzatore, eseguendo i processi cognitivi che sono necessari per effettuare tali scambi.

L'ammissione che l'utilizzatore non aveva (e non poteva essere richiesto di avere) alcuno stato cognitivo direttamente attinente agli specifici contratti stipulati dal suo AS (nessuna intenzione o volontà che tali contratti venissero formati, e nessuna conoscenza del loro contenuto e delle loro precondizioni) ci conduce ad un difficile dilemma riguardo ai contratti stipulati da AS (e in generale da sistemi automatici, tema sul quale si veda Allen and Widdison 1996). Dobbiamo decidere quale dei seguenti punti di vista adottare:

1. Questi contratti non sono accompagnati da alcun stato cognitivo giuridicamente rilevante. Si tratta di scambi senza accordo, per utilizzare la suggestiva espressione di Irti (1998), da considerarsi da una prospettiva puramente comportamentistica. Pertanto, aver adottato la decisione di stipulare un contratto, possedere l'informazione che sussistono circostanze di un certo tipo, credere che la controparte abbia certi stati cognitivi, dovrebbe essere irrilevante per quanto attiene agli effetti di questi contratti.

2. Questi contratti sono caratterizzati dagli stati cognitivi posseduti dagli AS che li stipulano. Pertanto il fatto che un AS avesse una certa intenzione, o certe credenze, almeno quando di ciò la controparte fosse consapevole, può incidere sugli effetti del contratto.

Si consideri ad esempio il seguente caso ipotetico. Si assuma che un AS abbia avuto l'incarico di vendere on line una partita di vecchi gioielli a seconda del loro peso, la loro età e il materiale di cui sono fatti. Si assuma che l'AS per stabilire il prezzo usi un database (preparato da un perito) dove esso può rinvenire una descrizione di tutti gli articoli da vendere. Si assuma che l'articolo n. 25 sia stato erroneamente classificato come anello d'argento con un copertura d'oro mentre si tratta di anello d'oro. Si assuma che

l'AS offra di vendere l'articolo 25 al prezzo di 20 euro, considerando che questo è il prezzo appropriato per un anello d'argento e che la controparte accetti. Si assuma altresì che dalla fotografia dell'anello visibile on line risulti chiaramente la dicitura « gold 18 K », e che la controparte accetti di acquistare l'anello al prezzo di 20 Euro, intuendo l'errore su cui si basa l'offerta dell'AS. Sarà il contratto annullabile, poiché la decisione di concluderlo si basava su un errore, noto alla controparte (la falsa credenza dell'AS che l'anello fosse d'argento) o sarà esso valido in quanto non vi è stato alcun errore (non potendo un AS trovarsi in uno stato cognitivo qualificabile come errore)? Se un contratto di questo tipo fosse stato concluso tra persone, esso sarebbe annullabile, in quanto viziato da errore essenziale e riconoscibile (anzi, riconosciuto) dalla controparte (art. 1428 c.c. ss). Che accade quando, come nel caso in esame, l'errore sia stato commesso da un AS? E che dire del caso in cui l'errore dell'AS sia stato indotto dalla controparte, che, ad esempio, ha modificato il database dell'AS?

L'impostazione comportamentistica, come abbiamo visto, ordina di astenersi dall'uso di nozioni cognitive nel considerare l'operato degli AS: ogni effetto giuridico deve essere direttamente collegato a comportamenti osservabili, e non a stati cognitivi che possano essere inferiti da comportamenti osservabili. Ciò che conta è solo il fatto che siano stati inviati certi messaggi, che hanno certi significati secondo le convenzioni sociali prevalenti. Nell'esempio che abbiamo appena considerato sopra, poiché erano stati inviati appropriati messaggi di offerta e di accettazione, dovremmo concludere a favore della validità del contratto. Come questo esempio mostra, l'impostazione comportamentistica, mentre può essere ragionevole in una certa misura (come quando le parti abbiano convenuto di attribuire certi effetti prestabiliti a certe azioni dei loro sistemi) può portare a risultati assurdi, e non può fornire la flessibilità di un'impostazione basata su nozioni intenzionali.

Il problema del modello comportamentistico è che è impossibile specificare in anticipo quale comportamento osservabile corrisponderà a certi stati cognitivi (ad esempio, alla credenza in certi fatti, o all'intenzione di produrre un certo risultato). Se noi collegiamo direttamente e necessariamente a certi comportamenti osservabili gli effetti giuridici usualmente giustificati (anche) dalla presenza di stati cognitivi, allora tali effetti si verificheranno talvolta anche se quegli stati cognitivi sono assenti, e essi potranno non verificarsi quando essi sono presenti. Tale impostazione si presta ad essere sfruttata opportunisticamente, come accade quando un AS sia indotto con l'inganno a inviare un certo messaggio, pur non avendo l'« intenzione » di compiere l'azione comunicativa corrispondente. Inoltre, una conseguenza spiacevole di questa impostazione è la necessità di duplicare ogni nozione giuridica, al fine di fornire equivalenti comportamentistici delle nozioni intenzionali usate per governare le interazioni umane. Dovremmo

quindi assumere che i contratti informatici siano un modo nuovo di creare relazioni giuridiche, diverso dai contratti tra esseri umani come sosteneva Giannantonio (1997).

La possibilità alternativa consiste nel prendere sul serio la prospettiva intenzionale, cioè nell'attribuire stati cognitivi agli AS (come ad altri sistemi informatici), e nel collegare a tali stati specifiche conseguenze giuridiche. La prospettiva intenzionale, quale unica attitudine a noi disponibile nell'accostarci a sistemi complessi ed autonomi, verrebbe ad avere un riconoscimento giuridico. Seguendo questa impostazione, la controparte di un AS potrebbe interpretare le dichiarazioni contrattuali dell'AS attribuendo allo stesso le corrispondenti intenzioni (l'intenzione di vendere l'anello al prezzo di 20 euro), e inoltre gli stati epistemici presupposti da tali intenzioni (la credenza che l'anello sia d'argento).

Di regola, il fatto che l'AS effettui una certa dichiarazione, in circostanze appropriate, sarà indicazione sufficiente circa gli stati intenzionali dell'AS, e il fatto che l'AS non abbia l'intenzione di effettuare tale dichiarazione sarà irrilevante, qualora la controparte ragionevolmente credesse che l'AS avesse tale intenzione (pertanto, nella grande maggioranza dei casi, l'indirizzo comportamentistico e quello intenzionale darebbero lo stesso risultato).

Tuttavia quando l'AS (pur inviando un certo messaggio) non avesse alcuna intenzione di stipulare un contratto e la controparte fosse consapevole di ciò (o dovesse esserlo, date le circostanze del caso), nessun contratto potrebbe dirsi concluso. Si consideri, ad esempio, l'ipotesi che l'AS abbia inviato una sequenza di parole (corrispondente ad un'offerta contrattuale) nel corso di esperimenti intesi a verificare la disponibilità e l'affidabilità di una linea di trasmissione.

Inoltre, i difetti nei processi cognitivi dell'AS, quali impedimenti nella formazione della volizione contrattuale, dovrebbero avere implicazioni simili a quelle dei tradizionali vizi del volere (errore, violenza, dolo). Allo stesso modo dovrebbero considerarsi le ipotesi nelle quali l'AS possieda conoscenze che (se possedute da un contraente umano) pregiudicherebbero l'efficacia del contratto (ad esempio, la conoscenza che la controparte non è proprietaria del bene mobile oggetto della compravendita).

Il fatto che il contenuto del contratto sia stato determinato (anche) da parte dell'AS, non esclude che i diritti e i doveri creati mediante il contratto ricadano sull'utilizzatore. Ciò è proprio quanto l'utilizzatore voleva, nel momento in cui ha delegato la formazione del contratto al suo AS. Pertanto, l'intenzione dell'utilizzatore (come riconoscibile alla controparte) di delegare la formazione del contratto alla cognizione dell'AS è la ragione per la quale il contratto vincola l'utilizzatore (è « non-ripudiabile » da quest'ultimo), benché l'utilizzatore non abbia voluto lo specifico contenuto del contratto concluso dal suo AS. I diritti e le obbligazioni stabi-

liti dal contratto ricadranno quindi sull'utilizzatore non perché egli abbia voluto creare quelle posizioni giuridiche, ma perché egli ha scelto di delegare al suo AS la formazione di contratti in suo nome. La cognizione dell'AS si aggiunge quindi alla cognizione dell'utilizzatore, secondo l'intenzione dell'utilizzatore, e dovrebbe essere trattata, in linea di principio, nello stesso modo. Ciò condurrebbe ad assimilare la situazione dell'utilizzatore di un AS alla situazione di chi affidi la conclusione di un contratto ad un rappresentante¹⁶. Ciò che le due situazioni hanno in comune, e che le distingue dalle ipotesi nelle quali si utilizzi un mezzo di trasmissione meccanico o umano, è la delegazione cognitiva, cioè la decisione di affidare la formazione del contenuto di un contratto e la decisione se concluderlo o meno (benché all'interno di obiettivi e vincoli prestabiliti) alla cognizione del proprio sostituto (umano o artificiale).

Questa prospettiva esclude che ogni determinazione dell'AS sia necessariamente (o debba essere) una determinazione del suo utilizzatore: quando sia necessario stabilire chi volesse che cosa, dobbiamo esaminare quali contenuti del contratto siano stati prestabiliti dall'utilizzatore, e quali siano stati determinati dall'AS. Di conseguenza quando si debba stabilire se la conclusione del contratto sia dovuta a dolo (di modo che il contratto debba essere annullato), circa gli elementi che sono stati determinati dall'utilizzatore si deve esaminare se l'utilizzatore sia stato ingannato, ma rispetto agli elementi determinati dall'AS, si deve esaminare se l'AS stesso sia stato indotto in errore. Per quando riguarda gli effetti dell'errore, si deve considerare però che l'errore incide sulla validità del contratto, solo se riconoscibile alla controparte. Ciò circoscrive gli effetti degli errori (false credenze o falsi stati epistemici), sia quando siano stati fatti dall'utilizzatore che quando siano stati compiuti da un AS. Se l'errore dell'AS non è riconoscibile alla controparte, la dichiarazione contrattuale fatta dall'AS (nell'ambito in cui l'AS appare ragionevolmente operare secondo il mandato dell'utilizzatore) continuerà a vincolare l'utilizzatore.

Sembra quindi che a questo riguardo il modello della rappresentanza e del mandato (art. 1387 ss. c.c.) possano fornire indicazioni adeguate. Infatti, secondo questo modello, le disfunzioni del processo deliberativo dell'AS (e conseguentemente i suoi stati mentali) incidono sulla validità del contratto. Questa tesi sembra essere compatibile con l'asserto seguente, incluso nel summenzionato documento Uncitral:

« il Gruppo di lavoro ha adottato l'opinione che, pur essendosi usata l'espressione « agente elettronico » per ragioni di conve-

¹⁶ Come è noto questa idea è stata avanzata da Borruso (1988) con riferimento ai contratti informatici in generale.

nienza, l'analogia tra un sistema automatico e un agente commerciale non sia appropriata. Pertanto, i principi generali dell'« agency law » (ad esempio i principi attinenti alla limitazione di responsabilità per il risultato del comportamento errato dell'AS) non possono essere usati in relazione alle operazioni di tali sistemi. Il Gruppo di lavoro ha reiterato la sua precedente convinzione che, per principio generale, la persona (fisica o giuridica) per conto della quale un computer è stato programmato dovrebbe essere il responsabile ultimo di ogni messaggio generato dalla macchina (A/CN.9/484, para. 107). Per regola generale, l'utilizzatore di uno strumento è responsabile per i risultati ottenuti mediante l'uso dello strumento poiché lo strumento non ha una volizione propria. Tuttavia, un « agente elettronico » per definizione è capace, entro i parametri della sua programmazione, di iniziare, rispondere o interagire con altre parti o con i loro agenti elettronici una volta che sia stato attivato da una parte, senza ulteriore attenzione di quella parte¹⁷ ».

In particolare, sembra da condividere l'opinione che sarebbe inappropriato trasferire automaticamente agli utilizzatori di AS ogni regola applicabile al rapporto tra rappresentato e rappresentante (o mandante e mandatario), poiché, in particolare, il rappresentante può rispondere con il proprio patrimonio. Invece, sembra errata la tesi della connessione tra responsabilità (soggezione a conseguenze giuridiche) dell'utilizzatore e assenza di volontà dello strumento utilizzato: come abbiamo osservato, quando l'utilizzatore sceglie di usare uno strumento cognitivo, lo si può ritenere responsabile anche per azioni risalenti alla determinazione (volontà) dello strumento (un ripensamento in questo senso sembra peraltro emergere dall'ultima frase della citazione appena riportata). Pertanto, il fatto che sia nel conferimento di rappresentanza (con mandato), sia nell'utilizzo di un AS nella contrattazione venga attuata una delegazione cognitiva può giustificare utili analogie, poiché le stesse ragioni possono consigliare le stesse soluzioni giuridiche.

¹⁷ The Working Group was of the view that, while the expression « electronic agent » had been used for purposes of convenience the analogy between an automated system and a sales agent was not appropriate. Thus, general principles of agency law (for example, principles involving limitation of liability as a result of the faulty behaviour of the agent) could not be used in connection with the operation of such systems. The Working Group reiterated its earlier understanding that, as a general principle, the person (whether a natural person or a legal entity) on whose

behalf a computer was programmed should ultimately be responsible for any message generated by the machine (A/CN.9/484, para. 107). As a general rule, the employer of a tool is responsible for the results obtained by the use of that tool since the tool has no independent volition of its own. However, an electronic agent, by definition, is capable, within the parameters of its programming, of initiating, responding or interacting with other parties or their electronic agents once it has been activated by a party without further attention of that party.

Si consideri ad esempio la questione del tempo della conclusione del contratto, rispetto agli AS mobili. Un AS mobile può operare a distanza senza interagire con l'utilizzatore e con il computer di quest'ultimo (ad esempio, un AS può risiedere su un *mobile device*, come un telefonino, e operare in rete anche quando il device è sconnesso dalle linee telefoniche). Se l'AS fosse solo un mezzo per la comunicazione tra l'utilizzatore e le controparti, i contratti stipulati mediante l'AS risulterebbero conclusi, secondo il nostro diritto, solo quando l'accettazione della controparte raggiungesse l'utilizzatore (o almeno il sistema informatico di quest'ultimo). Ciò può porre alcune difficoltà in alcune applicazioni degli AS. Si immagini, ad esempio un AS mobile che, introdottosi in un marketplace finanziario, inizia ad acquistare e vendere titoli. E ragionevole assumere che i contratti stipulati mediante l'AS siano conclusi solo quando l'accettazione della controparte raggiunge il computer dell'utilizzatore? Ciò impedirebbe all'AS di vendere quanto esso abbia acquistato, prima di comunicare l'acquisto all'utilizzatore (poiché la comunicazione all'utilizzatore è necessaria per completare lo scambio precedente), e pertanto impedirebbe all'AS di partecipare a scambi ravvicinati. Anche in questo caso il modello della rappresentanza sembra fornire indicazioni utili per accostarsi alla negoziazione basata sugli AS: il contratto stipulato mediante un rappresentante è di regola concluso quando l'accettazione raggiunge il rappresentante, soluzione estendibile anche agli AS.

6. CONCLUSIONE.

I temi generali che abbiamo esaminato nelle pagine precedenti non esauriscono le problematiche attinenti alla disciplina giuridica degli AS. L'utilizzo degli AS è infatti rilevante anche in numerosi ambiti specifici del diritto, come la protezione dei consumatori, la proprietà intellettuale, la protezione dei dati, il diritto all'informazione, le imprese virtuali, la risoluzione delle controversie on line¹⁸. Non esamineremo neppure il controverso problema della opportunità di attribuire personalità giuridica agli AS¹⁹ se non per ribadire l'irrelevanza di tale problema rispetto alla questione qui trattata, cioè quella del valore giuridico degli stati intenzionali posseduti dagli (o attribuiti agli) AS, e degli atti intesi ad esprimere tali stati. Il tema della personalità degli AS attiene infatti alla pos-

¹⁸ Per una raccolta di contributi in materia, cf. Cevenini (2002), per brevi rassegne, cf. Weitzenboeck (2001b), Sartor (2002a).

¹⁹ Rispetto al problema della persona-

lità degli agenti concordiamo con la soluzione negativa proposta dalla maggior parte degli studiosi. Per alcune brevi cenni e richiami bibliografici, *cf.* Sartor (2002a).

sibilità che gli AS abbiano diritti e doveri propri, e in particolare che essi siano titolari di un patrimonio distinto e separato rispetto a quello dei proprietari e utilizzatori. Noi ci siamo invece limitati a considerare le ipotesi nelle quali l'AS viene utilizzato per effettuare o ricevere dichiarazioni da cui dipendono effetti sulla sfera giuridica di altri soggetti.

La soluzione affermativa da noi raggiunta circa la possibilità di attribuire rilievo giuridico all'intenzionalità degli AS, in particolare negli ambiti della responsabilità civile e dei contratti, non impone quindi alcuna modifica nel diritto delle persone. Al fine di ottenere una disciplina giuridica adeguata agli AS (negli ambiti da noi considerati) bisognerebbe solo applicare analogicamente le norme in materia di responsabilità vicaria (combinandole con quelle attinenti alla responsabilità di proprietari e custodi) e in materia di sostituzione nell'attività giuridica (mandato e rappresentanza).

Infine, l'attribuzione stati intenzionali rileva non solo per gli agenti software ma anche per altri sistemi informatici, e potrebbe consentire in generale di trasporre, con gli adattamenti del caso, la disciplina che regola le relazioni umane anche ai rapporti con, e tra, entità cognitive artificiali²⁰.

RIFERIMENTI BIBLIOGRAFICI.

- T. ALLEN and R. WIDDISON (1996), *Can computers make contracts*, in *Harvard Journal of Law and Technology* 9, 25-52.
- R. BORRUSO (1988), *Computer e diritto. Problemi Giuridici dell'informatica*, Vol. 2, Milano, Giuffrè.
- C. CASTELFRANCHI and R. FALCONE (1998), *Towards a theory of delegation for agent-based systems*, in *Robotics and Autonomous Systems* 24, 141-157.
- C. CEVENINI, J. GELATI, and A. VILLECCO (2002). *Proceedings of LEA 2002. Workshop on the Law of Electronic Agents*, Bologna, CIRSFID.
- A. CICU (1901), *Gli automi nel diritto privato*, in *Il Filangeri* 8, 1-30.
- R. CLARIZIA (1985), *Informatica e conclusione del contratto*, Milano, Giuffrè.
- M. DAVIES (1998), *The philosophy of mind*, in A. GRAYLIN (Ed.), *Philosophy 1. A Guide Through the Subject*, pp. 250-335, Oxford, Oxford University Press.

²⁰ Per più ampie considerazioni su questo tema, e per alcuni riferimenti bibliografici, cf. Sartor (2002b).

- F. DE MIGLIO, T. ONIDA, F. ROMANO, and S. SANTORO (2002), *Electronic agents and the law of agency*, in *Proceedings of LEA 2002. Workshop on the Law of Electronic Agents*, pp. 23-32, Bologna, CIRSFID.
- D.C. DENNETT (1989), *The Intentional Stance*, Cambridge (Mass), MIT.
- D.C. DENNETT (1991), *Consciousness Explained*, Boston (Mass), Little Brown.
- D.C.C DENNETT (1997), *Kinds of Minds: Towards an Understanding of Consciousness*, New York, Basic Books.
- D.C. DENNETT and J.C. HAUGELAND (1987), *Intentionality*, in R. GREGORY (Ed.), *The Oxford Companion to the Mind*, pp. 383-386, Oxford, Oxford University Press.
- F. DRETSKE (1986), *Misrepresentation*, in R. BOGDAN (Ed.), *Belief: Form, Content and Function*, pp. 17-36, Oxford, Oxford University Press.
- G. FINOCCHIARO (2002), *The conclusion of the electronic contract through « software agents »: A false legal problem? brief considerations*, in *Proceedings of LEA 2002. Workshop on the Law of Electronic Agents*, pp. 75-80, Bologna, CIRSFID.
- E. GIANNANTONIO (1997), *Diritto dell'informatica* (Seconda ed.), Milano, Giuffrè.
- N. IRTI (1998), *Scambi senza accordo*, in *Rivista trimestrale di diritto e procedura civile*, 347-364.
- J. LEROUGE (2000), *The use of electronic agents questioned under contractual law: Suggested solutions on a european and american level*, in *The John Marshall Journal of Computer and Information Law* 18 (2), 430-00.
- J. PITT, A. MAMDANI, and P. CHARLTON (2001), *The open agent society and its enemies: A position statement a programme of research*, in *Telematics and Informatics* 18 (1), 67-87.
- S.J. RUSSELL and P. NORVIG (1995), *Artificial Intelligence. A Modern Approach*, Englewood Cliffs (NJ), Prentice Hall.
- G. SARTOR (2002a), *Gli agenti software: nuovi soggetti del ciberdiritto*, in *Contratto e impresa*, 57-91.
- G. SARTOR (2002b), *L'intenzionalità dei sistemi informatici e il diritto*, in *Rivista trimestrale di diritto e procedura civile*, In pubblicazione.
- G. WEISS (1999), *Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence*, Cambridge (Mass), MIT.
- E. WEITZENBOECK (2001a), *Electronic agents and the formation of contracts*, in *International Journal of Law and Information Technology* 9 (3), 204 - 234.
- E. WEITZENBOECK (2001b, 13 October 2001), *Electronics agents: some other legal issues*. ECLIP 2nd Summer School, Palme De Mallorca. Disponibile presso il sito: <http://www.eclip.org/summerschool/2nd/presentations>.

- E. WEITZENBOECK (2002), *Electronic agents and contract performance: Good faith and fair dealing*, in *Proceedings of LEA 2002. Workshop on the Law of Electronic Agents*, pp. 67-74, Bologna, CIRSIFID.
- M. WOOLDRIDGE (2000), *Reasoning about Rational Agents*, Cambridge (Mass), MIT.
- A. YIP and J. CUNNINGHAM (2002), *Some issues on agent ownership*, in *Proceedings of LEA 2002. Workshop on the Law of Electronic Agents*, pp. 13-22, Bologna, CIRSIFID.