ANÁLISE DOS CONCEITOS DE AUTONOMIA E RESPONSABILIDADE E O CONTEXTO DA AGÊNCIA ARTIFICIAL

MARÍLIA

2016

FRANCIELE DA SILVA LEAL

ANÁLISE DOS CONCEITOS DE AUTONOMIA E RESPONSABILIDADE E O CONTEXTO DA AGÊNCIA ARTIFICIAL

Dissertação de Mestrado apresentada junto ao Programa de Pós-Graduação em Filosofia da Universidade Estadual Paulista, Campus de Marília, como requisito para a obtenção do título de Mestre em Filosofia.

Área: Filosofia da Mente, Epistemologia e Lógica.

Linha de pesquisa: Filosofia da mente, Ciência Cognitiva e Semiótica

Orientadora: Profa. Dra. Mariana Claudia Broens

Agência financiadora: CAPES

MARÍLIA

2016

Leal, Franciele da Silva.
L435a Análise dos conceitos de autonomia e responsabilidade e o contexto da agência artificial / Franciele da Silva Leal. – Marília, 2016.
f.87 ; 30 cm.

Orientador: Mariana Claudia Broens.
Dissertação (Mestrado em Filosofia) – Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, 2016.
Bibliografia: f. 84-86

1. Inteligência artificial - Filosofia. 2. Inteligência artificial. 3. Robôs. 4. Tecnologia e ética. I. Título.

CDD 001.535

FRANCIELE DA SILVA LEAL

ANÁLISE DOS CONCEITOS DE AUTONOMIA E RESPONSABILIDADE E O CONTEXTO DA AGÊNCIA ARTIFICIAL

Dissertação apresentada junto ao Programa de Pós-Graduação em Filosofia, da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista Júlio de Mesquita Filho, Campus de Marília, como requisito parcial para a obtenção do título de mestre em Filosofia, sob a orientação da Profa. Dra. Mariana Claudia Broens.

Área de concentração: Filosofia da Mente, Epistemologia e Lógica.
Linha de Pesquisa: Filosofia da Mente, Ciência Cognitiva e Semiótica

Data de Exame da defesa: 30/09/2016

Membros da Banca Examinadora:

Titular 1 (orientadora): Profa. Dra. Mariana Claudia Broens (UNESP/Marília).
Titular 2: Prof. Dr. Marcos Antonio Alves (UNESP/Marília)
Titular 3: Prof. Dr. Anor Sganzerla (PUCPR)

Suplente interno: Edna Alves Souza (UNESP/Marília)
Suplente externo: Marco Aurélio Sousa Alves (UFSJ)

Dedico esse trabalho aos meus pais, Neusa e Francisco, e a minha orientadora, profª Mariana.

Agradecimentos

Agradeço primeiramente aos meus pais, pelo incentivo dado durante toda a minha vida, para que continuasse meus estudos. Aos meus irmãos, Fábio que leu carinhosamente o meu trabalho, e ao Gervásio. Ao meu companheiro, Ricardo, por estar sempre ao meu lado, me pondo para cima e me fazendo acreditar que posso mais do que imagino.

Também gostaria de agradecer à professora Mariana Claudia Broens, minha orientadora, que me auxiliou muito, querendo que eu aproveitasse cada segundo dentro do mestrado para absorver algum tipo de conhecimento, e não desistiu de mim, mesmo com todas as minhas limitações. Serei eternamente grata!

Agradeço também, a todas as amizades que construí em Marília e que desejo levar para sempre em meu coração. Em especial, a Amanda e a Iracelis que me acolheram em seus lares, e me fizeram sentir parte da família. A Nathália que contribuiu para o desenvolvimento do trabalho, com leituras e sugestões. A Silvia e a Josy pelas trocas de experiências e pelo incentivo para não desistir no decorrer do caminho. Ao Vini, pelas risadas, ao Paulo Martins e Paulo Uzai pelas conversas. E agradeço também a todos os colegas do GAEC (Grupo Acadêmico de Estudos Cognitivos) cujas discussões auxiliaram na elaboração deste trabalho.

Sou muito grata a todos os professores do programa pelas aulas ministradas. Em especial ao professor Marcos Alves, que me incentivou a ingressar no mestrado ainda na graduação, não só com palavras, mas com sua postura em sala de aula. E à professora Maria Eunice Gonzalez, detentora de muito conhecimento, não só acadêmico, mas conhecimentos de vida que me auxiliou a refletir sobre muitas ações cotidianas de que lembrarei a vida toda.

Agradeço a todos os funcionários da Secretaria de Pós-Graduação da FFC _ UNESP e em especial a Edna Bonini de Souza, secretária do Departamento de Filosofia, por sua eficiência, carinho e disposição para sempre ajudar.

Por fim, agradeço à CAPES pelo apoio financeiro.

Ninguém vence sozinha... OBRIGADA A TODOS E TODAS!

Uma população estática poderia em um dado momento dizer "basta!", mas uma população crescente precisa dizer "mais!"

(Hans Jonas, 2013, p.166).

LEAL, F.S. Análise dos conceitos de autonomia e responsabilidade e o contexto da agência artificial. Dissertação (mestrado em Filosofia). – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2016.

Resumo

Neste trabalho nos propomos a desenvolver uma análise crítica dos conceitos de agência e responsabilidade no contexto tecnológico contemporâneo em que são produzidos sistemas artificiais ditos autônomos. Tendo esse objetivo em foco, apresentamos primeiramente teorias da ação e problematizamos as noções de evento causal e ação causal, objeto e agente, visando clarificar a noção de agência. Analisamos, em seguida, a noção de ação responsável proposta por Hans Jonas uma vez que ele propõe uma noção de responsabilidade considerando o desenvolvimento tecnológico contemporâneo. Em especial, apresentamos e discutimos a noção de heurística do temor proposta por Jonas (2004) cujo objetivo é criar cenários possíveis que antecipem implicações a longo prazo do uso de novas tecnologias e auxiliem a informar a sociedade sobre as possíveis consequências de seu uso para as novas gerações.Por fim, tratamos mais especificamente da legitimidade da atribuição de agência e responsabilidade a sistemas artificiais, especialmente a robôs, na perspectiva da Filosofia e da Robótica, a partir de teses defendidas por Hans Jonas (2004, 2013) e Willem Haselager (2005). Em seus textos, ambos autores discutem a possibilidade de atribuir agência e responsabilidade a sistemas artificiais capazes de aprender e tomar decisões que ultrapassem os limites de sua programação inicial. Concluímos que parece problemático tanto atribuir agência a sistemas artificiais autônomos no mesmo sentido em que se considera que seres humanos são agentes quanto reduzir tais sistemas a meros objetos causalmente determinados.

Palavras-chave: Agência; Autonomia; Responsabilidade; Sistemas artificiais autônomos; Heurística do temor.

LEAL, F.S.An analysis of the concepts of autonomy and responsibility in the context and artificial agency.Dissertação (mestrado em Filosofia). – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2016.

Abstract

In this work, we propose to develop a critical analysis of the concepts of agency and responsibility in the contemporary technological context in which artificial systems said to be autonomous are produced. Aiming on this goal, we present, firstly, some theories of action, and we problematize the notions of causal event and causal action, object and agent, aiming on clarifying the notion of agency. Following, we analyze the notion of responsible action proposed by Hans Jonas, once he proposes a notion of responsibility considering the contemporary technological development. In particular, we present and discuss the heuristic notion of fear proposed by Jonas (2004) whose goal is to create possible scenarios which may anticipate long-term implications of the use of new technologies and may help to inform society about the possible consequences of their use for the new generations. Finally, we deal more specifically with the legitimacy of the attribution of agency and responsibility to artificial systems, specially robots, from the perspective of Philosophy and Robotics, from theses defended by Hans Jonas (2004, 2013) and Willem Haselager (2005). In their texts, both authors discuss the possibility of attributing agency and responsibility to artificial systems capable of learning and making decisions which surpass the limits of their initial programming. We conclude that it seems problematic both to attribute agency to autonomous artificial systems in the same sense in which human beings are considered to be agents, and to reduce such systems to mere causally determined objects.

Key-words: Agency; Autonomy; Responsibility; Autonomous artificial systems; Heuristics of fear.

SUMÁRIO

Introdução

O objetivo geral desta dissertação é investigar os conceitos de agência e responsabilidade e suas relações no âmbito da teoria da ação, e buscar discutir estes conceitos no contexto tecnológico de sistemas artificiais ditos autônomos.

Para iniciar nossa análise, consideremos os seguintes exemplos: um motorista entra em seu carro e o carro não aceita ser ligado porque o motorista apresenta sinais de embriaguez; um avião bélico que voa sem controlador humano remoto e pode alterar sua rota inicial por ter detectado um alvo mais relevante para ser bombardeado; um robô com a função de auxiliar os seres humanos em seu cotidiano como, por exemplo, ir buscar um copo de água, jogar cartas, servir chá e fazer a limpeza de locais de trabalho ou residências. Não estamos tratando de situações hipotéticas, pois já existem carros com controle antiálcool em países como Reino Unido, Estados Unidos e Canadá (Noorman, 2014), bombardeios autônomos como o *Taranis*(Prado, 2015), e robôs como o PR2 que podem desenvolver diversas funções para assistir pessoas com necessidades especiais (Garage, 2015). Assim, podemos constatar que sistemas artificiais estão presentes no cotidiano do ser humano contemporâneo, presença essa que tende a aumentar com o desenvolvimento de novas tecnologias informacionais com um maior grau de autonomia e, na mesma medida, de Interação/interferência/ direcionamento da conduta humana.

Considerando as possibilidades dessas novas tecnologias, neste trabalho investigamos se, e, em caso afirmativo, em que sentido a autonomia pode ser atribuída legitimamente a sistemas artificiais, o que também implica a análise da legitimidade de atribuição de responsabilidade ou corresponsabilidade a tais sistemas, seus idealizadores e fabricantes. Esta investigação se revela relevante até porque, frequentemente, os criadores de novas tecnologias parecem perder o controle sobre seu desenvolvimento e uso no momento em que apresentam suas criações à comunidade, a partir das quais são implementadas aplicações, muitas delas sequer antevistas pelos idealizadores das tecnologias originais[1].

[1]Lembremos do caso do uso de aviões convencionais para fins bélicos durante a primeira grande guerra, finalidade que não havia sido antevista pelo seu inventor, Alberto Santos Dumont, o que teria levado Santos Dumont ao suicídio.

Tendo em vista que atualmente já há projetos de sistemas artificiais em desenvolvimento que talvez possam ser considerados autônomos em algum sentido, consideramos relevante repensar conceitos éticos, tais como os de agência, autonomia e responsabilidade e suas possíveis correlações. Em síntese, entendemos que o sentido de tais conceitos precisa ser atualizado de acordo com o novo contexto contemporâneo em que a noção clássica de agência (no sentido básico de possuir a capacidade de agir por si mesmo), que remete a habilidade antes considerada apenas atribuível a seres humanos, pode vir a se estender a outros seres, naturais e artificiais.

Desse modo, com o objetivo, já enunciado, de analisar os conceitos de autonomia, agência e responsabilidade no novo contexto tecnológico, consideramos, que teses apresentadas por Hans Jonas (2006) em torno do conceito de responsabilidade associado ao desenvolvimento e uso de tecnologias podem nos auxiliar a refletir sobre problemas éticos relacionados à possibilidade de sistemas artificiais agirem por si mesmos, isto é, poderem ser considerados autônomos em algum sentido. Assim, tendo em vista tal objetivo geral, dividimos nosso trabalho em três capítulos.

No primeiro capítulo, ***Análise de Teorias da Ação e suas relações com o conceito de agência,*** procuramos analisar teses centrais de algumas teorias que buscam elucidar a natureza da ação. O conceito de ação é relevante em nossa pesquisa, pois precisamos delimitar o que pode ser considerado uma ação, sobretudo, quais características são necessárias para considerar um evento do mundo como uma ação autônoma praticada por um ser dotado de agência e, nessa medida, passível de gerar vínculos de responsabilidade sobre seus efeitos. Tendo em vista esse propósito, neste capítulo apresentamos teses sobre a natureza da ação segundo Alicia Juarrero (1999), Roberto Casati e Achille Varzi (2015) para iniciarmos a discussão da diferenciação entre agente causal e evento causal, isto é, quando uma alteração de estados de coisas no mundo pode ser considerada como resultante de um agente. Juarrero (1999) relaciona ação, evento e agente e ressalta que o evento não deve ser o único a causar um comportamento no agente, mas o agente também deve agir causalmente. Também introduzimos a noção de agência e de autonomia coletiva e compartilhada apresentada por Markus Schlosser (2015), o que nos dá subsídios para contrapormos aos conceitos de responsabilidade coletiva. E finalmente fazemos uma breve introdução à noção de

ação proposta por Hans Jonas (2004), cujo conceito de responsabilidade e relações com a tecnologia serão discutidos no segundo capítulo deste trabalho.

No segundo capítulo, intitulado ***O princípio responsabilidade: reflexões de Hans Jonas acerca da tecnologia,*** analisamos a concatenação dos conceitos de autonomia e responsabilidade. Tendo em vista nosso objetivo geral de investigar tais conceitos no contexto de sistemas artificiais considerados autônomos, encontramos nos três primeiros capítulos da obra *Técnica, medicina e ética: sobre a prática do princípio responsabilidade* de Hans Jonas uma discussão relevante sobre a noção de responsabilidade no contexto do desenvolvimento tecnológico, especialmente o contemporâneo, para ser desenvolvida em nosso trabalho. Além disso, procuramos ressaltar as discussões levantadas por Jonas (2006) em torno às consequências que as novas tecnologias podem trazer a longo prazo, sobretudo as tecnologias de uso bélico. A reflexão sobre as tecnologias bélicas foi uma das principais causas da proposta de uma "heurística do temor", instrumento de análise ética que propicia a reflexão sobre diferentes cenários futuros e as possíveis consequências a longo prazo resultantes das tecnologias nesses cenários, procurando antecipar as piores consequências que poderiam ser geradas a partir da ação presente, de modo a evitar as implicações nefastas de novas tecnologias para as gerações futuras. As reflexões de Jonas sobre responsabilidade e tecnologia são importantes para refletir sobre a situação contemporânea de novos sistemas artificiais que não se reduzem a instrumentos ou ferramentas tecnológicas, mas eles incorporam em algum grau a capacidade de decidir, como é discutido no terceiro capítulo deste trabalho.

Assim, no terceiro capítulo ***A relação entre autonomia, agência e corporeidade: corpo biológico e corpo artificial,*** investigamos em que sentido sistemas naturais e artificiais têm, ou não, propriedades em comum, especificamente se gozam de fato de agência e autonomia em algum sentido semelhante ou aproximado. Desse modo, neste terceiro capítulo será apresentada uma distinção inicial entre sistemas artificiais automáticos e autônomos proposta pela *Royal Academy of Engineering* (2009), segundo a qual estes últimos se caracterizam por terem a capacidade de aprender com ações passadas e tomar decisões não previstas em sua programação inicial. Partindo desta caracterização, analisamos e discutimos teses sobre a natureza da autonomia na perspectiva da Filosofia e na perspectiva dos estudos da Robótica apresentadas por

Willem Haselager (2005) como complementares e não necessariamente dicotômicas. Adotando a perspectiva da Cognição Situada e Incorporada, Haselager (2005) argumenta que, embora não seja ainda possível atribuir autonomia plena a sistemas artificiais por não possuírem capacidades homeostáticas (relacionadas à automanutenção equilibrada de seus estados internos), nem por isso se deve recusar a sistemas artificiais, especialmente os robóticos, a posse de algum grau de autonomia se forem capazes de aprender e tomar decisões. Procuramos ressaltar que a importância atribuída à corporeidade aproxima teses defendidas por Jonas (2004, 2013) e Haselager (2005) em torno do papel que o corpo desempenha na ação autônoma e na agência.

Entendemos que, a partir da proposta de considerar distintos níveis de autonomia na ação, tal como apresentado por Harry Frankfurt (1978), que pressupõe não ser necessária uma intenção para realizar uma ação em grau mais "fraco", é legítimo considerar que sistemas autônomos artificiais possuam agência, mesmo que em um grau mais fraco, sendo assim possível discutir os graus de autonomia e agência de sistemas artificiais. No entanto, uma questão que procuraremos discutir diz respeito à pertinência de atribuição de responsabilidade a tais sistemas artificiais autônomos em algum grau, pois tal adjudicação não parece, pelo menos a primeira vista, adequada devido a sua origem com funcionalidades predeterminada por sua programação.

Assim sendo, procuraremos neste trabalho investigar e discutir a legitimidade de considerar sistemas artificiais autônomos em algum grau como detentores de agência e potencialmente corresponsáveis, na mesma proporção de sua autonomia, por possíveis implicações de suas ações no mundo.

CAPÍTULO 1
Análise de Teorias da Ação e suas relações com o conceito de *agência*

Apresentação

Neste capítulo apresentamos e discutimos o conceito de agente, algumas abordagens sobre sua possível natureza e suas principais características. Tal discussão se faz necessária para alcançarmos o objetivo geral deste trabalho que consiste em investigar o conceito de responsabilidade e autonomia em relação aos sistemas artificiais, ainda que essas noções de autonomia e responsabilidade sejam delineadas em um sentido fraco. Tal necessidade resulta de que a imputação de responsabilidade e autonomia é feita no âmbito da ação.

Tendo em vista o objetivo deste capítulo, nos dedicaremos primeiramente a apresentar diferentes perspectivas sobre a natureza da ação. Para isso, nos embasaremos especialmente em teses apresentadas por Alicia Juarrero (1999), George Wilson e Samuel Shpall (2012), Roberto Casati e Achille Varzi (2015), Harry Frankfurt (1978) eHans Jonas (2004).

1.1- A distinção entre agentes causais e eventos causais

Inicialmente, consideramos que, como ressaltam Juarrero (1999) e Wilson e Shpall (2012), a teoria da ação é uma área da filosofia que busca primeiramente identificar o que caracteriza uma ação considerada voluntária (embora o conceito de vontade permaneça obscuro) e a diferencia de um movimento reflexo de organismos. Para fazer essa investigação, a teoria da ação se utiliza dos conhecimentos de diversas áreas tais como Metafísica, Epistemologia, Psicologia, Filosofia da Mente, Inteligência Artificial e Neurociências, entre outras (Juarrero, 1999, p. 2). Ao analisar o conceito de ação com o auxílio dessas áreas do saber, outros conceitos como comportamento voluntário, intenção, vontade, consciência, causalidade, agente/agência, evento, causação mental, entre outros, se tornam objetos de estudo essenciais para que a teoria da ação alcance seu objetivo.

Sem pretender analisar todos esses conceitos, nesta dissertação daremos ênfase a alguns deles e a algumas teorias da ação, que passamos a analisar. De acordo com Juarrero (1999), Wilson e Shpall (2012), cabe iniciar a investigação da natureza da ação a partir da distinção entre eventos causais e agentes causais, na medida em que os

eventos causais supõem determinações enquanto agentes causais supõem algum grau, menor ou maior, de possibilidade de escolha. Apontam Wilson e Shpall (2012, s/p) que:

> Para produzir uma questão central sobre a natureza da ação, comumente é invocada uma distinção intuitiva entre as coisas que só acontecem com as pessoas - os eventos a que elas estão submetidas - e as várias coisas que elas genuinamente provocam. Estes últimos, o que as pessoas fazem, são os atos ou ações do agente, e o suposto problema sobre a natureza da ação é: o que distingue uma ação de um mero acontecimento ou ocorrência?[2]

Segundo Casati e Varzi (2015), um acontecimento pode ser observado na variedade de elementos dispostos no mundo. O nosso planeta não é formado apenas por diferentes seres vivos, objetos físicos e elementos abstratos, mas também pela variedade de coisas que acontecem ou são ocasionadas por esses elementos, como por exemplo, chegadas e partidas, festas, danças, raios, trovões e explosões. A observação desses diversos acontecimentos instigou a filosofia nas últimas décadas a investigar, com mais dedicação, a noção de *evento*. Essa pesquisa resultou em implicações para áreas como a Ciência Cognitiva, a Filosofia da Mente, a Linguística, entre outras. Thomas F. Shipley (2008, s/p) complementa:

> O conjunto de eventos não é o conjunto de todas as coisas. No entanto, eventos são uma parte tão grande do "todo" que pode ser útil considerar o que não é um evento. Eventos são coisas que acontecem; eventos requerem uma referência a uma localidade no tempo (mas não necessariamente a um ponto no tempo). Algo que existe fora de tal referência temporal - vamos chamá-lo *objeto* - não é um evento. Objetos físicos (por exemplo, maçãs, montanhas, nuvens) não são claramente eventos e objetos psicológicos (por exemplo, ideias, conceitos, metas) tampouco são eventos. Um objeto isolado não é um evento; eventos ocorrem quando objetos se alteram ou interagem. Uma maçã não é um evento, mas a queda de uma maçã o é. Da mesma forma, a ideia da gravidade não é um evento, mas ter a ideia foi um evento.

Levando em consideração as noções de evento, objeto e agente previamente apresentadas, buscaremos investigar qual dessas definições seria mais adequada para caracterizar os sistemas artificiais contemporâneos, até porque tais noções não são tão claras e simples como podem aparentar. Casati e Varzi (2015), por exemplo, afirmam

[2] It has been common to motivate a central question about the nature of action by invoking an intuitive distinction between the things that merely happen to people — the events they undergo — and the various things they genuinely do. The latter events, the doings, are the acts or actions of the agent, and the problem about the nature of action is supposed to be: what distinguishes an action from a mere happening or occurrence?

que, embora a definição de evento apresentada acima possa parecer clara a primeira vista, ela está longe de sê-lo. A expressão "coisas que acontecem" para caracterizar a noção de evento, por exemplo, é problemática e obscura devido à própria vagueza do termo 'acontecer' e à frequente circularidade de tal caracterização ("O que é um evento? Algo que acontece; e o que acontece? Acontecem eventos").

A partir dessa observação, os autores se dispõem a apresentar e problematizar as relações do termo evento com outros termos, tais como evento e objeto, evento e fato, evento e propriedade e evento e tempo. No presente estudo focalizaremos a relação entre evento e objeto, pois acreditamos ser relevante para nos auxiliar a contrastar a noção de evento e ação, até porque a possibilidade de haver ações produzidas por sistemas artificiais torna problemática a caracterização de agente enquanto *sujeito* intencional, pois, se houver agentes artificiais, teríamos que admitir a possibilidade de existirem objetos intencionais ou teríamos que admitir que pode haver sujeitos artificiais.

Para iniciar o contraste entre os conceitos de evento e objeto, Casati e Varzi(2015) ressaltam as diferenças entre os termos "objeto" e "acontecimento", embora ambos estejam relacionados. A fim de ir mais a fundo nessa diferenciação, os autores apresentam quatro distinções:

> Se uma distinção metafísica entre objetos e eventos é aceita, em seguida surge uma questão quanto à relação entre entidades nas duas categorias. Os objetos são os atores principais em eventos; eventos sem objetos são incomuns. Mas também são incomuns objetos sem eventos; eventos compõem a existência dos objetos. Em um modo radical, no entanto, pode-se pensar as entidades em uma categoria sendo metafisicamente dependentes das entidades da outra categoria (CASATI E VARZI ,2015, s/p).[3]

A primeira diferenciação entre objeto e evento apresentada por(Hacker Cresswell apud*:* Casati e Varzi 2015) consiste em afirmar que os objetos físicos "existem" enquanto os eventos "acontecem", ou seja, podemos afirmar que "existe uma pedra em cima da montanha", a "pedra" é considerada um objeto. E na proposição "choveu essa manhã", o verbo "chover" é considerado um evento. A partir desta

[3] If a metaphysical distinction between objects and events is granted, then a question arises as to the relation between entities in the two categories. Objects are prime actors in events; objectless events are uncommon. But so are eventless objects; events make up the lives of objects. In a radical mood, however, one can think of the entities in one category as being metaphysically dependent on entities in the other.

primeira diferenciação é possível afirmar que os sistemas artificiais são objetos, pois também estão extremamente relacionados com eventos. Por exemplo, a *Tapia robot*[4], (dispositivo digital)poderia ser considerada um objeto que existe, enquanto na preposição " o sistema falhou" pode ser considerado um evento.

O segundo aspecto diferencial é baseado nos autores Quinton (1979) e Hacker (1982b) e afirmam que ambos, eventos e objetos, se relacionam com o espaço e o tempo de maneira distinta. Os objetos têm limites espaciais mais nítidos e limites temporais mais vagos, enquanto no evento acontece o inverso, os limites espaciais são mais vagos enquanto os limites temporais são relativamente nítidos. Por exemplo, se um agente *X* pergunta ao agente *Y* "onde está a pedra?" à qual se referiu na conversa que tiveram ontem, é possível o agente *X* levar o agente *Y* até a montanha e mostrar a pedra, sendo relativamente "fácil" limitar o espaço e mostrar a localização da pedra. Todavia se o agente *Y* perguntar sobre a idade da pedra e há quanto tempo ela se encontra naquela localização específica, a resposta seria mais complexa, pois embora existam estudos que possam fazer uma estimativa sobre o tempo de duração da pedra e desde quando se encontra naquele lugar, não seria uma informação tão segura quanto a primeira. Em contraposição, se o agente *Y* pergunta onde está (espaço) a chuva da manhã, não seria mais possível localizá-la, embora pudesse mostrar a rua e a terra molhadas, essas não seriam a chuva propriamente dita, aquele evento em que gotículas de água se precipitam sobre o solo. Em compensação, se o agente *Y* perguntar sobre o horário em que a chuva ocorreu na cidade naquele dia, a resposta, consideravelmente precisa, seria que choveu entre as 8h e as10h da manhã.

O terceiro aspecto diferencial é apresentado por (Dretske apud: Casati e Varzi 2015). Para ele, os objetos se movem enquanto os eventos não se movem. Como por exemplo, a pedra pode se mover, embora necessite de uma força externa, de modo que um agente *X* pode levar a pedra para sua casa, ou uma tempestade fazer com que ela role montanha abaixo. Entretanto quando nos referimos à chuva embora em um sentido popular possamos afirmar que "a chuva está caindo", o que está efetivamente caindo são gotículas de água; não é possível um agente ou força externa mover a chuva, o agente *X*

[4]*Talk Robot Companion*. Criada para distração e para dar assistência as pessoas em seu dia a dia, ela faz compras online, informa a previsão do tempo, faz ligações, avisa datas importantes, como um aniversário, entre outras funções. Mais informações em http://www.japantrendshop.com/tapia-ai-robot-companion-p-3460.html

não pode levar o evento “chuva” para casa, apenas parte da água que caiu quando dizemos “está chovendo”.

O quarto e último aspecto diferencial apontado por (Johnson ;Mellor; Simons apud: Casati e Varzi 2015) consiste na alegação de que o objeto é considerado contínuo no tempo, ele está presente em todos os momentos enquanto ele existir. Em contrapartida, os eventos são esporádicos e estão presentes em tempos específicos. É o caso da pedra que, por exemplo, pode permanecer durante décadas em uma montanha e a chuva que pode aparecer na montanha em momentos isolados como, por exemplo, em alguns dias do verão.

Entretanto, a última distinção segundo Casati e Varzi (2015) gerou várias controversas, pois há outros autores, entre eles Goodman (1951), que aceitam que evento e objeto encontram-se no mesmo nível, ou pertenceriam à mesma categoria, pois os objetos seriam considerados eventos mais “monótonos” e eventos, poderiam ser considerados objetos mais “instáveis”.

Cabe apontar que, no contexto teórico que distingue *evento* e *objeto*, objetos não têm poder causal a não ser quando postos em movimento por forças físicas, como é o caso das gotículas de água de chuva, em si mesmas objetos, movendo-se segundo o ciclo hidrológico. Em tal contexto teórico, apenas ***agentes*** (organismos dotados de *agência*) teriam o poder de iniciar uma cadeia causal de eventos no mundo, uma vez que a ocorrência de eventos estaria previamente determinada pelas leis físicas. No entanto, se a distinção entre evento e objeto não é tão claramente delineável, como indaga Goodman (1951), será clara a distinção entre agente causal e evento causal? Então enfrentamos mais um obstáculo para clarificarmos a natureza dos sistemas artificiais contemporâneos.

Uma vez que, devido a sua natureza física, distinta da natureza dos organismos biológicos por sua composição e carência de uma história adaptativa, sistemas artificiais, como *Tapia* e outros robôs, são considerados objetos, eles seriam ontologicamente equivalentes às gotas de água do exemplo acima. No entanto, tal equivalência ontológica parece problemática no contexto tecnológico contemporâneo, como iremos discutir posteriormente.

Após esta reflexão inicial sobre a relação entre os conceitos de evento e objeto, nos propomos a investigar mais especificamente a noção de ***agente causal*** e suas

principais características. Para esse fim, nos baseamos nas discussões das perspectivas causais e suas respectivas críticas apresentadas por Alicia Juarrero em sua obra *Dynamics in action: Intentional Behavior as a Complex System*. Nesta obra, Juarrero (1999) analisa a distinção entre agente causal e evento causal. Para alcançar esse fim, a autora recorre, inicialmente, à ideia de Roderick Chisholm (1964) de que, para haver a distinção entre agentes e eventos causais, é necessário acreditar que alguns eventos são causados por não eventos, nesse caso por *agentes*.

O conceito de agente não tem uma caracterização única. Ele é caracterizado de acordo com a perspectiva teórica em que é analisado. Juarrero (1999) mostra isso ao longo do segundo capítulo de sua obra, no qual ela apresenta e problematiza algumas perspectivas da teoria causal da ação, na medida em que cada uma delas pressupõe uma determinada concepção de agente. Em uma perspectiva mais tradicional, por exemplo, para ser considerado um agente é necessário ter intenção, propósito; em outra perspectiva, a agência é concebida em níveis, fracos ou fortes, sendo que quando se trata dos níveis mais fracos, por exemplo, não se pressupõe a intenção para caracterizar a agência. Desse modo, para entender melhor as concepções de agente e evento e suas relações, propomos investigar, na próxima seção, os conceitos de agente causal e ação intencional e se sistemas artificiais podem ser considerados em algum sentido não trivial agentes causais intencionais.

1.2- Estudo crítico da abordagem causal da ação

De acordo com Juarrero (1999), alguns teóricos da ação defendem que não são apenas eventos que causam ou determinam o comportamento do agente, mas que este pode iniciar novas cadeias causais, não sendo, ou pelo menos não inteiramente, determinado por leis físicas. Mas frequentemente enfrentamos dificuldades de identificar um agente causal ou em que circunstâncias um agente inicia um encadeamento causal de eventos no mundo como resultado de sua autonomia, isto é, de suas próprias regras de conduta.

Iniciamos, então, com a reflexão acerca da causação na visão da Alicia Juarrero (1999). A autora inicialmente descarta a noção de causa de si (*self-cause),* ou auto causação, que pressupõe uma causa que não é resultado de eventos anteriores. Como a

própria autora afirma, há certo consenso entre os filósofos em descartar a noção da causa de si:

> Nada pode causar ou mover a si mesmo; nenhuma dependência ambiental é permitida. Uma vez que a filosofia moderna descartou a noção de causa final, não houve modo de conceber uma relação interna entre o mundo externo e o conteúdo de uma intenção como atrator e guia do comportamento voluntário. Uma vez que a filosofia moderna descartou a noção de causa formal, foi assumido que todas as causas são eventos que ocorrem operando somente como causas eficientes [...] (JUARRERO, 1999, p. 24).[5]

Ao negar a causa de si faz-se necessário explicar a relação de ação causal entre o agente e o evento, pois, segundo Juarrero (1999, p.26): "[...] para uma relação ser propriamente causal, os agentes devem diferenciar-se do comportamento que eles (poderosamente) causam. Nada pode ser causa de si mesmo"[6]. A autora afirma que, de modo geral, entre os filósofos tem sido aceito que as intenções desejos e crenças ativam a musculatura do corpo do agente a fim de movimentá-lo. O que não tem sido satisfatoriamente explicado é como tais intenções, desejos e crenças conseguem ativar a musculatura do corpo ou ter poder causal sobre ele.

Diante deste quase consenso sobre o papel das crenças e desejos como causas da ação, são apresentadas algumas teorias para refletir sobre a natureza causal da ação. A primeira apresentada por,(Arthur Danto apud: Juarrero, 1999) consiste na diferenciação entre ações básicas e ações não básicas. A ação básica é caracterizada por ser uma ação realizada diretamente pelo agente, como por exemplo, mover a mão, ele move a mão e nada mais. Em contraposição, uma ação não-básica pode ser apresentada por exemplo, por um agente que move a mão para apertar um interruptor para acender a luz. Essa ação constitui uma ação não básica, pois não se pode fazer qualquer uma dessas ações diretamente, ou seja, "para *acender* a luz é necessário *mover* a mão e *apertar* o interruptor", dessa forma afirma-se que as ações não básicas são ações realizadas através de outras ações (básicas).

Contudo, Juarrero (1999) ressalta que essa diferenciação causou vários desconfortos e indagações. A autora apresenta a seguinte hipótese para ilustrar melhor

---

[5] Nothing can cause or move itself; no environmental embeddedness is allowed. Once modern philosophy discarded the notion of final cause, there was no way to conceive of an internal relation between the outside world and the content of an intention as attractor and guide of voluntary behavior. Once modern philosophy discarded the notion of formal cause, all causes were assumed to be occurrent events operating only as efficient cause […].

[6] For a relationship to be properly causal agents must be other than the behavior they (powerfully) cause. Nothing can cause itself.

esses desconfortos aos quais se refere. Suponhamos que em uma situação corriqueira o agente *X* chega a sua casa e acende a luz. Para isso ele move o braço e aperta o interruptor, todavia o agente não esperava que com aquela ação ele pudesse alertar a um ladrão que estava em sua casa sobre sua chegada. Pergunta Juarrero: Quais desses eventos são ações do agente *X?* Frisando que o objetivo do agente *X* era acender a lâmpada e não alertar o ladrão.

Em uma possível resposta,(Chisholmapud:Juarrero,1999) afirma que uma ação básica pressupõe a intenção do agente em realizar a ação. No entanto, as teorias causais da ação são vulneráveis às objeções baseadas em cadeias causais irregulares (*causal chain objections*), ou seja, têm dificuldades em definir uma ação como causada intencionalmente pelo agente em alguns casos mais específicos, como no clássico exemplo do tio e sobrinho, que apresentaremos em seguida. Essa dificuldade se dá devido à noção de que o agente não pretendia realizar a ação *X* para alcançar o objetivo *Y*, ou que não desejava que a ação *X* tivesse um resultado inesperado.

O exemplo do tio e do sobrinho ao qual nos referimos e que é utilizado por Juarrero (1999) é explicado da seguinte maneira: Um agente *Y* percebe que para ter acesso à herança, ele necessita matar o tio. Nervoso com essa ideia, quando ele vai à casa de seu tio para praticar o homicídio, o nervosismo decorrente da antecipação do ato de matar o tio o leva a atropelar uma pessoa que, coincidentemente, era seu tio. Seria a morte do tio uma ação intencional do sobrinho? Deveria o sobrinho ser responsabilizado por homicídio culposo (acidental) ou doloso (intencional)?

Esse exemplo abre margens a muitas interpretações, o sobrinho poderia de fato ser considerado culpado, afinal, ele queria matar o tio e acabou matando-o. Outros poderiam afirmar que ele queria matar o tio, mas desejar realizar uma ação não tem o mesmo peso de uma ação concretizada e, assim sendo, o homicídio foi acidental, mesmo que o nervosismo gerador do acidente tenha resultado da antecipação do ato a ser cometido. Assim sendo, o sobrinho não pode ser responsabilizado por algo que ele não fez, apenas desejou fazer, mesmo porque, no momento em que sobrinho fosse executar a ação, ele poderia se arrepender e não matar o tio. No Brasil, juridicamente falando, se a morte do tio foi resultante de um acidente não intencional, o sobrinho possivelmente responderia por homicídio culposo. Juarrero (1999, p. 28) propõe que se considerem as causas como um evento ocorrente, e sendo ocorrente o desejo ou a

intenção não efetivam a ação, assim "... também é possível que fatores acidentais externos comprometam a cadeia causal e façam do comportamento resultante uma não-ação" [7].

Aproveitando o exemplo do indivíduo que acende a luz e sem saber alerta um ladrão sobre sua presença, propomos uma reflexão sobre o comercial da Tapia em que uma mulher aparece limpando um guarda-roupa e de repente começa a cantarolar uma música, no comercial *Tapia* prontamente reconhece a melodia e busca na internet a música que está sendo cantarolada e a põe para tocar. A mulher se mostra surpresa com a performance do robô, mas satisfeita. Entretanto, a surpresa gerada por performances de robôs com finalidades bélicas é obviamente de outra natureza. E o caso do robô Taranis, o qual, segundo a revista *Business Insider*[8], é uma aeronave não tripulada capaz de efetuar bombardeios autonomamente, sem controle remoto humano, que tem causado muitos questionamentos, inclusive de renomados pesquisadores como Stephen Hawking. O maior receio se deve a não ficar claro o limite do poder de decisão de tais sistemas para utilizar armamentos contra seres humanos.

Nessa perspectiva, é pertinente aprofundarmos a noção de agência tendo em vista que, segundo Schlosser (2015), o conceito de agência se refere à manifestação ou ao exercício da capacidade de agir de um agente. Nesse aspecto podemos investigar a possibilidade de atribuir agência a um sistema artificial. O autor argumenta que na perspectiva da teoria da ação tradicional é necessário a utilização dos conceitos de intencionalidade, causalidade, estados e eventos mentais para explicar o que pode ser considerado uma ação.

> Em um sentido bastante amplo, há agência virtualmente em toda parte. Sempre que as entidades estabelecem relações causais pode ser dito que atuam umas sobre as outras e interagem umas com as outras, provocando alterações entre si. Neste sentido muito amplo, é possível identificar os agentes e a agência (*agency)* e os passivos e a passividade (*patiency)*, praticamente em toda parte. Comumente, porém, o termo agência é usado em um sentido muito mais restrito para denotar o exercício de ações intencionais.[9](Schlosser, 2015, s/p.)

[7] It is also possible for extraneous, accidental factors to compromise the causal chain and make the resulting behavior non action.

[8] http://www.businessinsider.com/british-taranis-drone-first-autonomous-weapon-2015-9

[9] In a very broad sense, agency is virtually everywhere. Whenever entities enter into causal relationships, they can be said to act on each other and interact with each other, bringing about changes in each other. In this very broad sense, it is possible to identify agents and agency, and patients and patiency, virtually everywhere. Usually, though, the term 'agency' is used in a much narrower sense to denote the performance of intentional actions.

Assim, o próprio Schlosser (2015) ressalta que as teorias alternativas que sustentam que as concepções causais tradicionais da ação, não bastam para explicar a agência. Essas perspectivas alternativas propõem que tal conceito deve ser desenvolvido sem referência a causalidade eficiente, estados e eventos mentais. A discussão sobre a teoria causal da ação levanta muitos questionamentos, como sugere Juarrero (1999), pois para vários autores ela é insuficiente para explicar a distinção entre um agente e um evento causal. Desse modo, acreditamos que as críticas à teoria causal da ação apresentadas por Frankfurt (1978) sejam muito pertinentes para o nosso trabalho. Assim, embora a teoria causal da natureza da ação seja a justificação mais utilizada para fazer a distinção entre um evento e um agente, Frankfurt (1978) discorda das teorias causais que argumentam que a diferença essencial entre eventos com poder causal e agentes está presente nas causas que precederam a ação, de forma que um movimento corporal será definido como ação somente se for verificado que houve antecedentes de certo tipo.

A crítica que Frankfurt (1978) desenvolve contra a teoria causal se refere à natureza da ação, que, para ele, não pode ser explicada através de causas antecedentes de uma história causal particular, embora não negue que ações possam ter causas. Segundo Frankfurt (1978):

> Ao afirmar que a diferença essencial entre ações e meros acontecimentos reside em suas histórias prévias, as teorias causais supõem que ações e meros acontecimentos, em si mesmos, essencialmente não diferem. Essas teorias sustentam que as sequências causais que produzem ações são necessariamente de um tipo diferente daquelas que produzem meros acontecimentos, mas que os efeitos produzidos pelas sequências de ambos os tipos são inerentemente indistinguíveis. Portanto, eles estão comprometidos com a suposição de que uma pessoa que sabe estar no meio da prática de uma ação não pôde ter alcançado esse conhecimento de alguma consciência do que está acontecendo, mas, ao invés disso, deve ter alcançado esse conhecimento de seu entendimento de como o que está acontecendo foi causado por certas condições prévias. É parte integrante da abordagem causal considerar as ações e meros acontecimentos como sendo diferentes, não por algo que existe ou acontece no momento em que os eventos ocorrem, mas por algo bastante extrínseco a eles: a diferença com um outro conjunto de eventos em um momento anterior. Isto é o que faz que as teorias causais sejam implausíveis. (FRANKFURT, 1978, p. 157)[10].

[10] In asserting that the essential difference between actions and mere happenings lies in their prior causal histories, causal theories imply that actions and mere happenings do not differ essentially in themselves at all. These theories hold that the causal sequences producing actions are necessarily of a different type than those producing mere happenings, but that the effects produced by sequences of the two types are inherently indistinguishable. They are therefore committed to supposing that a person who knows he is in

Frankfurt (1978) afirma que, nesse tipo de análise da ação, a atenção se volta para longe do evento propriamente dito, no qual de fato estaria situada a causa, além de que a ação é analisada longe do tempo em que efetivamente ocorre. Após esses primeiros comentários críticos à teoria causal da ação, Frankfurt (1978) discute alguns aspectos das concepções de ação de David Pears que consideramos importantes para nosso trabalho.

Primeiramente, Frankfurt (1978) parte do pressuposto de que, o que é relevante é a ação em si, e não suas supostas causas antecedentes, enquanto a teoria causal da ação acredita que haja uma sequência de causas que produzem ações necessariamente distintas daquelas que produzem o mero movimento. Ele apresenta a seguinte linha de raciocínio.

> Durante o tempo em que a pessoa está realizando uma ação, ela está necessariamente em contato com os movimentos de seu corpo de um certo modo, enquanto que a pessoa não está necessariamente em contato com eles desse modo quando os movimentos de seu corpo ocorrem sem que ela os faça(FRANKFURT, 1978, p. 157)[11].

A partir disto, Frankfurt (1978) aponta que a teoria causal da ação não consegue dar conta dessa diferença de interação com os movimentos do próprio corpo quando é a pessoa que os faz e quando outra pessoa ou objeto os provoca devido ao foco exclusivo nas condições iniciais do movimento do corpo ou da ação.

Ao se utilizar da análise sobre a classificação da ação de Pears, Frankfurt (1978) ressalta que não se consegue distinguir com clareza quando se trata de uma ação ou de um movimento corpóreo. Mas Frankfurt (1978, p. 158, grifo nosso) complementa que "[...] disso não se segue, porém, que o único modo de descobrir se uma pessoa está agindo é considerar o que aconteceu antes de seus movimentos começarem, isto é, as causas de que se originaram. De fato, o estado de coisas durante a ocorrência dos

---

the midst of performing an action cannot have derived this knowledge from any awareness of what is currently happening, but that he must have derived it instead from his understanding of how what is happening was caused to happen by certain earlier conditions. It is integral to the causal approach to regard actions and mere happenings as being differentiated by nothing that exists or that is going on at the time those events occur, but by something quite extrinsic to them a difference at an earlier time among another set of events entirely. This is what makes causal theories implausible.

[11] During the time a person is performing an action he is necessarily in touch with the movements of his body in a certain way, whereas he is necessarily not in touch with them in that way when movements of hisbody are occurring without his making them.

movimentos é muito mais pertinente"[12]. Analisar se a ação está ocorrendo sob a orientação do agente é um fator determinante para apontar se o agente está ou não realizando a ação.

Todavia, Frankfurt (1978) discorda de David Pears no que tange a classificar a ação, o movimento provocado por outrem ou um espasmo muscular pela complexidade do movimento. Por exemplo, os movimentos dos dedos de um pianista tocando uma música não são espasmos musculares ou movimentos provocados por outrem, e sim ações muito complexas. Entretanto uma pessoa tendo uma crise epilética parece também conter movimentos bastante complexos, porém eles não são considerados como ações praticadas pelo indivíduo.

Dessa forma, Frankfurt (1978) propõe que a análise da complexidade do movimento corporal só faz sentido enquanto o movimento estiver em curso e estiver sob a orientação do agente, deixando de lado características causais anteriores. O autor ainda ressalta que o agente só estará executando uma ação se, e somente se, tal movimento estiver sob sua orientação, independentemente de que esse movimento "[...] forneça os antecedentes causais – na forma de crenças, desejos, intenções, decisões, volições ou outro qualquer – de que o movimento foi resultado" (Frankfurt, 1978, p.159)[13].

Ainda no que diz respeito às características da ação, Frankfurt (1978) alerta que é necessário entender a concepção do termo *intencional* utilizada no contexto em que ele apresenta sua teoria de *movimentos intencionais*. "Vamos empregar o termo "intencional" para nos referir a movimentos orientados, nos quais a orientação é fornecida pelo agente" (Frankfurt, 1978, p. 159)[14]. Essa expressão não deve ser confundida com a expressão *ação intencional*, pois, na perspectiva dele, se se trata de uma ação, ela deve necessariamente ser intencional: a expressão *ação intencional* constituiria, para ele, um pleonasmo, de modo que não se pode afirmar que a ação *X* é um movimento intencional. Mas haveria um sentido não trivial de ação intencional, que

[12] That the only way to discover whether or not a person is acting is by considering what was going on *before* his movements began that is, by considering the causes from which they originated. In fact, the state of affairs while the movements are occurring is far more pertinent.

[13] Provided the antecedent causes in the form of beliefs, desires, intentions, decisions, volitions, or whatever from which the movement has resulted.

[14] Let us employ the term "intentional" for referring to instances of purposive movement in which the guidance is provided by the agent.

segundo Frankfurt se deve preservar, em que as ações são praticadas deliberadamente de maneira autoconsciente. Acreditamos que o conceito de ação intencional proposto pelo autor nos ajuda a investigar se os movimentos e interações ambientais de sistemas artificiais podem ser considerados intencionais em algum sentido relevante.

A partir dessa perspectiva, o autor enumera dois problemas que os defensores dessa teoria da ação intencional devem enfrentar. O primeiro problema consiste em explicar a noção de comportamento orientado pelo agente. O segundo problema consiste em distinguir quando um comportamento é orientado efetivamente pelo agente ou resulta de processos fisiológicos, como quando a pupila se dilata em função da luz (p. 159). No caso do primeiro problema, aponta Frankfurt, trata-se de investigar as condições em que um movimento pode ser considerado proposital; no caso do segundo problema, por sua vez, trata-se de investigar as condições em que uma ação pode ser considerada intencional no sentido não trivial.

A fim de ilustrar essa discussão, Frankfurt (1978) apresenta um exemplo. O exemplo descreve todos os movimentos necessários para ligar um carro, desde pôr o cinto de segurança, colocar a chave na ignição, pisar na embreagem e colocar o câmbio em ponto neutro para enfim ligar a chave na ignição e dar a partida. Ele apresenta esse exemplo com o intuito de afirmar que nós não temos controle de nossos movimentos corpóreos no mesmo sentido em que o motorista tem controle sobre o carro, (ou o "capitão de seu navio", conforme a metáfora clássica!), pois não é possível descrever a totalidades dos movimentos que estariam sob a orientação do agente, correndo o risco de tratar-se quase de uma regressão ao infinito. Para Frankfurt (1978), afirmar que nossos movimentos sejam propositais não é o efeito de algo que fazemos, mas se trata antes de uma característica sistêmica do agente. Para uma melhor compreensão dessa tese, o autor nos traz uma definição de sua concepção de comportamento proposital que ocorre "... quando o seu curso está sujeito a ajustes que compensem os efeitos das forças que de outra forma poderiam alterá-lo e quando a ocorrência destes ajustes não é explicável por aquilo que explica o estado de coisas que os provoca" (Frankfurt, 1978, p.160)[15].

[15]Behavior is purposive when its course is subject to adjustments which compensate for the effects of forces which would otherwise interfere with the course of the behavior, and when the occurrence of these adjustments is not explainable by what explains the state of affairs that elicits them.

Nota-se que a questão do movimento corporal, ou seja, a presença de um corpo físico capaz de movimentos orientados, é um fator de peso nas análises para investigar a natureza da ação. Seria o "corpo" de um agente artificial capaz de instanciar ações intencionais? Aprofundaremos essa questão no terceiro capítulo.

Após apresentarmos a crítica de Frankfurt (1978) à teoria causal da ação e brevemente apresentar sua concepção sistêmica de ação, que posteriormente retomaremos, trataremos de outra vertente não causal da ação apresentada por Wilson e Shpall (2012) e Schlosser (2015).

1.3- Ação intencional e agência

Como vimos, Alicia Juarrero (1999) também estuda a natureza da ação intencional em um sentido não trivial a partir de uma perspectiva crítica da abordagem causal. Ela inicia sua análise sobre a natureza da ação retomando a pergunta clássica: Qual é a diferença entre um piscar de olhos (*wink*) e um movimento involuntário de nossa pálpebra (*blink*)? A questão levanta problemas quanto à noção de comportamentos ditos voluntários ou intencionais e os comportamentos considerados reflexos.

A autora apresenta o seguinte exemplo para mostrar a importância da diferenciação entre ações intencionais ou não. Um médico aplicou uma dose considerável de analgésico em seu paciente com dor, que posteriormente o levou a um estado de coma seguido de morte. Como categorizar o ato do médico? Trata-se de um homicídio? E, em caso afirmativo, trata-se de um homicídio doloso, isto é, proposital, ou um homicídio culposo, isto é, por negligência, imprudência ou imperícia? Nesse contexto se a *intenção* do médico foi aliviar a dor e se a dosagem administrada se encontra dentro dos parâmetros técnicos para casos semelhantes, sua ação não pode ser considerada como um homicídio seja doloso ou culposo.

No mesmo sentido, Wilson e Shpall (2012) nos fornecem outro exemplo sobre a natureza da ação intencional, ainda em um sentido não trivial. É comum que em determinadas situações, para afirmar ou negar algo, o agente *X* movimente a cabeça em um gesto de confirmação ou negação para o agente *Y* (o que seria uma ação *ativa*). Ou o agente *X* pode não ter movido a cabeça de maneira intencional, outro evento pode ter provocado o movimento da cabeça deste agente (podemos denominar nesse caso como

ação *passiva*) como, por exemplo, ter movido a cabeça para espantar um inseto de sua orelha. No entanto, as duas ações podem ser consideradas minimamente ativas. Então, como distingui-las?

Em resposta a essas indagações recorremos à discussão que Schlosser (2015) apresenta sobre concepções e teorias da ação e de agência. Ele ressalta que, em um sentido mais restrito, agênciapode ser entendida como o exercício de ações intencionais, no sentido não trivial. Mas que a mesma ação pode ser descrita de muitas formas e, algumas delas, podem não ressaltar aspectos intencionais. Schlosser (2015) apresenta o exemplo a que já nos referimos sobre o alerta (não intencional) a um ladrão feito pelo dono da casa ao chegar e acender uma luz (ação intencional).

Nessa perspectiva, uma sequência de movimentos pode envolver mais de uma "ação", mas só será considerada de fato uma ação se algum desses movimentos puder ser descrito como intencional. O segundo ponto é sobre a proposta da estreita relação entre intencionalidade e ter uma razão para agir (*acting for areason*) dada pela perspectiva tradicional. Nessa perspectiva se entende que para agir intencionalmente é necessário agir por alguma razão, ou seja, um agir decorrente de processos considerados racionais (o que leva, frequentemente, a limitar a atribuição de agir intencionalmente apenas a seres humanos dotados de capacidades cognitivas consideradas de alto nível).

Contudo, já nos referimos à dificuldade de distinguir uma ação intencional de uma ação não intencional por não haver um consenso geral do que seja intenção, dando margem a diferentes interpretações e usos do termo, como por exemplo, os diferentes sentidos que a noção de intenção apresenta nas sentenças: *"ter intenções para o futuro"*, *"agir intencionalmente"* e *"agir com certa intenção".*

Além dessa dificuldade, Wilson e Shpall (2012) apontam que o verbo "fazer" também é problemático e muito abrangente para distinguir a ação de um movimento provocado no agente ou de um mero espasmo muscular. Tanto é assim que é possível classificar o "tossir" e "espirrar", considerados movimentos reflexos, como uma forma de "fazer" do agente, embora na maioria das vezes seja entendido que se trata de um fazer "passivo", embora isto pareça paradoxal. Poder-se-ia dizer que esse tipo de fazer não é relevante para a teoria da ação, porém é extremamente difícil explicar então em qual sentido o *fazer* poderia ser tratado adequadamente em uma teoria da ação.

Podemos encontrar também outras definições do fazer ativo na teoria da ação. Frankfurt (1978), por exemplo, propõe que o fazer ativo pode ser dividido em distintos níveis. Os níveis mais complexos seriam próprios das ações humanas, enquanto animais não humanos apresentariam um nível menos complexo. Embora o autor afirme que animais não humanos tenham o controle direto de suas ações, direcionadas ou não, e sobre seus movimentos com propósitos e objetivos, de forma que teriam um fazer ativo, mas o têm em um nível mais fraco devido a não possuírem as ferramentas de autoconsciência que, supostamente, apenas os seres humanos possuem[16]. Isto porque, segundo Frankfurt, os seres humanos planejam, estabelecem metas com base em avaliações mais amplas de suas possibilidades futuras, suas opções e oportunidades. Para Frankfurt, o diferencial em relação aos animais não humanos é a capacidade humana de reflexão e a consciência imediata da realização da atividade *X* em função de determinado fim preestabelecido.

Schlosser (2015) concorda com Frankfurt e considera que a teoria tradicional, que pressupõe que somente os seres humanos são capazes de agir intencionalmente, em razão de suas capacidades representacionais e linguísticas, estaria sendo muito exigente. Todavia se aceitarmos a teoria da ação intencional em um viés menos radical e mais amplo, poderíamos propor níveis de agência, de modo a incluir as ações de organismos não humanos que revelam propósitos.

Além disso, Schlosser (2015) ressalta que estudos desenvolvidos nas áreas da Filosofia da Mente, Ciência Cognitiva, Robótica e na teoria dos sistemas dinâmicos revelam a possibilidade de condutas coletivas de sistemas artificiais, em si mesmos bastante simples, puderem ser consideradas como intencionais, mesmo que em um sentido mais restrito (Brooks; Beerapud: Schlosser 2015). O foco desta concepção é sua estratégia de agir em rede, o que permite aos agentes interagirem com outros agentes, sendo que essa interação os habilita a responder a exigências de uma situação complexa de uma forma hábil, sem muito esforço, deliberação consciente, raciocínio ou planejamento.

[16] Este argumento de Frankfurt, embora ainda muito comum, reflete apenas a tese não comprovada de que as capacidades cognitivas humanas seriam excepcionais e restritas à espécie humana. Atualmente essa tese, frequentemente revestida de preconceitos, está sendo contestada em várias áreas do conhecimento: Etologia, Psicologia Evolucionária, Inteligência Artificial, entre outras.

Segundo ressalta Schlosser (2015), vários autores entendem que a capacidade de agência não está restrita aos seres capazes de produzir representações mentais ou de terem habilidades linguísticas sofisticadas. Segundo Schlosser (2015) esta concepção é adotada por psicólogos da percepção direta como Gibson (1986) e filósofos da mente como Chemero (2009), Silberstein e Chemero (2011), Hutto e Myin (2014). Esse argumento é uma espécie de generalização do argumento anterior, de que existe agência que resulta diretamente da percepção de informação significativa no ambiente dos agentes que direciona suas possibilidades de ação. Contudo, ele questiona como explicar a nossa capacidade de deliberar sobre o futuro e formular raciocínios abstratos sem assumir a mediação de representações mentais, questões estas muito difíceis e que permanecem em aberto.

A partir da discussão das teses retomadas por Schlosser (2015), constatamos quão complexas e problemáticas são as noções de ação e de agência e quão abrangente pode ser a utilização do conceito de agência. Segundo Schlosser (2015), a concepção tradicional de agência (intencional no sentido representacional e linguístico) é de onde derivam as demais concepções de agência, inclusive aquelas que procuram não incluir o conceito de intencionalidade como um atributo necessário. Essas noções de agência vão de níveis mais "elaborados" tais como agência autocontrolada (*self-controlled*), livre (*free agency*) e autônoma (*autonomous*) até os níveis menos elaborados que não exigem uma linguagem mental para a atribuição da agência. Nos capítulos dois e três desta dissertação investigaremos a legitimidade de atribuição de agência a sistemas artificiais que gozam de algum grau de autonomia.

Além da agência intencional, encontramos novas perspectivas da noção de agência, as quais incluem concepções de agência compartilhada, de agência coletiva e a possibilidade de agência artificial, como no exemplo apresentado por (Chemero apud: Schlosser 2015).

No que tange à diferenciação entre agência coletiva e agência compartilhada Schlosser (2015) esclarece que a agência coletiva ocorre quando dois agentes desenvolvem uma ação, como por exemplo, cantar uma música ou transportar algum móvel; enquanto a agência compartilhada ocorreria quando dois ou mais indivíduos agem como um grupo de acordo com certos princípios. Umas das questões propostas a

essa noção pela teoria tradicional da *agência* é saber se faz sentido atribuir desejos, estados e eventos mentais para grupo de indivíduos.

Ainda no que diz respeito à agência coletiva, este conceito é muito pertinente para este trabalho no que tange à discussão sobre a noção de responsabilidade coletiva, até mesmo para problematizar o tópico da agência artificial, por exemplo, seria legítimo responsabilizar um ser humano e um sistema artificial por um resultado mal sucedido de uma ação praticada conjuntamente?

Enquanto alguns estudiosos se preocupam em resolver a questão da existência da responsabilidade coletiva, segundo Marion Smiley (2011) outro grupo levanta questões já subentendendo sua existência: os grupos têm de cumprir as mesmas condições rigorosas de responsabilidade moral que os indivíduos cumprem? Quais as vantagens e desvantagens de manter determinados tipos de grupos, por exemplo, os estados-nação, culturas, grupos étnicos, como moralmente responsáveis na prática? Esta pergunta se torna especialmente relevante quando se trata de atribuir ou não responsabilidades. Lembremos que, finda a segunda guerra mundial, foi atribuída à nação alemã responsabilidade pelos atos brutais praticados em campos de concentração, embora tenha havido alemães individualmente contrários à prática de tais atos.

A filósofa Hannah Arendt dedica um subtítulo de sua obra *Responsabilidade e julgamento* para analisar a responsabilidade coletiva. Nessa perspectiva, a reponsabilidade coletiva se refere a atos que não praticamos diretamente (responsabilidade vicária[17]), e que por isso, não se poderia culpar um indivíduo por algo em que ele não teve participação ativa: "quando somos todos culpados a culpa não é de ninguém [...] Assim como temos compaixão para com aqueles que sofrem (e então nasce a solidariedade), declarar que somos todos culpados é uma declaração de solidariedade para com os 'malfeitores'" (Arendt, 2004, p. 214). A filósofa explica quais condições são, segundo ela, necessárias para que uma ação possa ser de responsabilidade coletiva:

> Diria que duas condições têm de estar presentes para a responsabilidade coletiva: devo ser considerado responsável por algo que não fiz, e a razão para a minha responsabilidade deve ser o fato de eu pertencer a um grupo (coletivo), que nenhum ato meu pode dissolver, isto é, o meu pertencer ao

[17] "Responsabilidade vicária" é uma expressão vinda do direito romano que designa a responsabilidade de um indivíduo pelas implicações de atos praticados por outrem.

> grupo é completamente diferente de uma parceria de negócios que posso dissolver quando eu quiser... Esse tipo de responsabilidade, na minha opinião, é sempre política, quer apareça na forma mais antiga em que toda uma comunidade assume a reponsabilidade por qualquer ato de qualquer de seus membros, quer no caso de uma comunidade ser considerada responsável pelo que foi feito em seu nome. (Arendt, 2004, p. 216).

Nessa perspectiva, para Hannah Arendt (2004), por exemplo, todo governo assume a responsabilidade pelos atos de seus predecessores (assim como toda a nação). Desse modo, quando dizemos que somos responsáveis pelos "pecados de nossos pais", por exemplo, assim dizemos por que ainda desfrutamos das recompensas de tais "pecados". Todavia não podemos ser considerados culpados nem juridicamente e nem moralmente e tampouco recebemos méritos pelos atos que nossos pais praticaram. Para "escapar" da responsabilidade coletiva seria necessário abandonar a comunidade, mas os seres humanos não podem viver sem alguma comunidade.

Outra visão de responsabilidade coletiva nos é apresentada por Marion Smiley (2011), a qual afirma que a responsabilidade coletiva exige, não apenas uma reflexão sobre as ações e intenções coletivas, mas também a análise do conceito de uma "mente coletiva", detentora de crenças e desejos coletivos, o que tem provado ser um dos maiores desafios para quem deseja manter uma noção de responsabilidade coletiva. Por exemplo, uma das condições para que a responsabilidade coletiva possa ser atribuída a certo grupo, consistiria em verificar o compartilhamento de crenças e desejos expressos nas ações coletivas e a adesão dos membros do grupo a tais crenças.

Ao longo da história, a reflexão filosófica sobre a responsabilidade moral segundo Marion Smiley (2011) focaliza a atribuição de responsabilidade somente ao indivíduo. Por outro lado, o fator tecnológico tem influenciado as reflexões sobre responsabilidades compartilhadas, e são levantadas algumas questões, como as que seguem: Quem é responsável pelas informações publicadas na Internet? Quem é responsável quando os registros eletrônicos são perdidos ou quando eles contêm erros? Em que medida e em que período de tempo são os desenvolvedores de tecnologias de informação responsáveis por eventuais consequências desagradáveis dos seus produtos? E, na medida que os produtos das novas tecnologias informacionais se tornam mais complexos e podem se comportar com certo grau, crescente, de autonomia, deveriam os

seus desenvolvedores e produtores ainda serem responsabilizados pelo comportamento desses sistemas artificiais autônomos?[18]

Não há ainda respostas satisfatórias para as questões apresentadas acima; todavia, Marion Smiley (2011) se reporta à noção de sistemas *sociotécnicos*. Essa noção se refere às situações em que são distribuídas tarefas entre humanos e sistemas artificiais. Quando ocorre um evento inesperado de consequências negativas, quanto mais complexo for o sistema, mais difícil é identificar de quem ou do que foi a responsabilidade pela falha.

Nesse sentido, os estudos sobre agência artificial, segundo Schlosser (2015), propõem investigar a possibilidade de atribuir agência a sistemas artificiais detentores de algum grau de autonomia. Schlosser (2015) argumenta que se adotarmos a visão tradicional de agência poderia haver filósofos, como Daniel Dennet, por exemplo, que aceitariam atribuir estados mentais a sistemas artificiais. Em contraposição, os filósofos que consideram que os seres humanos são detentores de uma intencionalidade originária, isto é, são autônomos para iniciar novas cadeias causais no mundo, como é o caso de John Searle, não aceitariam que estados mentais e representacionais possam ser atribuídos a sistemas artificiais, mesmo detentores de algum grau de autonomia. Por fim, Schlosser (2015) apresenta a noção de agência mínima, ou seja, da capacidade de agir autonomamente sem a necessidade de possuir estados mentais intencionais. Esta concepção minimalista de agência será novamente tratada no Capítulo 3.

Para os propósitos deste trabalho, adotamos a concepção sistêmica de agência proposta por Frankfurt (1978), pois consideramos pertinente a crítica à teoria causal da ação e sua tentativa infrutífera de determinar os antecedentes causais para distinguir ação de movimento. Desse modo, adotamos a concepção de que agência supõe a capacidade de ajustar o curso da ação, mesmo que em graus distintos, para compensar efeitos de forças que poderiam alterar tal curso, sendo que "... a ocorrência destes ajustes não é explicável por aquilo que explica o estado de coisas que os provoca"(Frankfurt, 1978, p.160). Também adotamos neste trabalho a noção de responsabilidade compartilhada entre agentes que atuam em função de aderirem a

[18] No texto *A lógica da ação*, Krister Segerberg*et al* (2013) discorrem sobre a descrição de agentes artificiais e os definem da seguinte maneira "... os agentes são entidades de *software* que exibem formas de inteligência / racionalidade e autonomia. Eles são capazes de tomar a iniciativa e tomar decisões por conta própria, sem o controle direto de um controlador humano" (Segerberg*et al*, 2013, s/p).

sistemas de crenças específicos, como sugere Marion Smiley (2011), e que podem constituir sistemas sóciotécnicos, constituídos pela integração de agentes naturais e artificiais, propiciando assim algum grau de agência para sistemas artificiais. Entretanto investigaremos se tal nível pode ser considerado mais ou menos elaborado como propõe Schlosser (2015). Para tanto teremos que investigar o conceito de responsabilidade e autonomia, uma vez que nosso este capítulo nos apontou para uma possível agência de sistemas artificiais compartilhada com seres humanos, assim questionamentos como já feitos neste capítulo surgem e são primordiais a serem explicitados.

Depois de apresentar uma discussão sobre diferentes teorias da ação e suas contribuições para compreender a noção de agência, adotando a concepção sistêmica proposta por Frankfurt para os fins deste trabalho, no próximo capítulo analisar a teoria ética que gira em torno do conceito de responsabilidade proposta pelo filósofo Hans Jonas. Esta escolha se mostra particularmente adequada à discussão sobre a responsabilidade compartilhada por sistemas sóciotécnicos, no sentido proposto por Smiley (2011), levando em conta sistemas artificiais dotados de algum grau de autonomia, isto é, capazes de auto ajustar o curso de suas ações. Além disso, Jonas também disponibiliza uma reflexão ética sobre a tecnologia e nos inspira a questionar até que ponto é aconselhável, e sob quais condições, deveríamos investir e desenvolver certos tipos de sistemas artificiais capazes de realizar performances altamente sofisticadas.

CAPÍTULO 2
O princípio responsabilidade: reflexões de Hans Jonas acerca da tecnologia

Apresentação

Este capítulo tem como objetivo refletir sobre a relação entre os conceitos de responsabilidade e tecnologia sob o viés sugerido pelo filósofo contemporâneo Hans Jonas (1903-1993). Tendo em vista nosso principal objetivo de analisar a possibilidade de atribuir agência e responsabilidade a sistemas artificiais atualmente produzidos capazes de auto ajustar o curso de suas ações, encontramos na proposta de Jonas (2006) uma reflexão contemporânea sobre as relações entre ética e tecnologia. Uma importante contribuição do filósofo para a discussão dessas relações é a distinção que propõe entre a dinâmica formal da tecnologia enquanto um empreendimento coletivo com suas próprias "leis de movimento" e o conteúdo substancial da tecnologia, isto é, as novas possibilidades de ação que a tecnologia oferece para o ser humano. Considerando esta distinção, concluiremos o capítulo discutindo a noção de responsabilidade proposta por Jonas, buscando refletir sobre implicações éticas do desenvolvimento tecnológico e das alegadas necessidades e vantagens de algumas dessas tecnologias.

2.1- A dinâmica formal da tecnologia contemporânea segundo Hans Jonas

Nesta seção objetivamos analisar a reflexão de Jonas sobre a tecnologia contemporânea. Tal objetivo específico se faz pertinente tendo em vista o objetivo geral deste trabalho de investigar os conceitos de autonomia e responsabilidade no contexto dos sistemas artificiais. Assim, iniciamos este capítulo com uma diferenciação conceitual importante para tal reflexão, pois aponta alguns aspectos que distinguem a "técnica" antiga da "tecnologia" contemporânea, cuja dinâmica de desenvolvimento e aplicação subverte a relação clássica entre meios e fins no âmbito da ação humana, antes presente no desenvolvimento técnico. Aponta Hans Jonas que:

> Se o conceito de "técnica", *grosso modo*, denomina o uso de ferramentas e dispositivos artificiais para atividades da vida, junto com sua invenção originária, fabricação repetitiva, contínua melhora e ocasionalmente também adição ao arsenal existente, tão tranquila descrição serve para a maior parte da técnica ao longo da história da humanidade (a qual tem a mesma idade que ela), mas não para a moderna tecnologia. (Jonas, 2013, p. 27).

Ou, com outras palavras, a técnica pode ser definida como "[...] o acúmulo de instrumentos de decifração e domínio sobre a realidade, o que faz dele um 'meio' e não um 'fim' em si mesmo". (Sganzerla, 2011, p.116). Em contraste, o termo tecnologia deve ser entendido como se referindo as ferramentas e procedimentos que estão relacionados ao desenvolvimento de uma civilização ou a uma versão mais apurada e desenvolvida de uma técnica, como por exemplo, a agricultura e a tecnologia da informação.

No passado, as ferramentas e seu uso eram caracterizados por serem constantes e bastante equilibrados. Ainda que existissem "revoluções", como por exemplo, a agrícola e a metalúrgica, é possível afirmar que ambas não foram promovidas propositalmente enquanto tais, pois ocorreram muito lentamente, a ponto de serem denominadas "revoluções" a partir de uma "contração temporal da retrospectiva histórica" (Hans Jonas, 2013, p. 27). Entretanto, a produção do primeiro carro bélico, como a biga egípcia, é apontada pelo autor como uma novidade repentina que, assim como outras técnicas bélicas, em vez de se estenderem pelo mundo, foram inicialmente monopolizadas e guardadas pelas sociedades inventoras.

Sobre o progresso das técnicas no período anterior à modernidade o autor enfatiza que "as melhoras foram esporádicas e não planejadas e o progresso, portanto – se é que se produzia – consistia em acréscimos insignificantes a um nível geralmente alto (do progresso moderno) que ainda hoje desperta nossa admiração [...]" (Jonas, 2013, p. 28-29). Em contraposição, a técnica moderna se mostra de modo inverso ao da técnica pré-moderna e Hans Jonas (2013) apresenta quatro justificativas para essa afirmação.

A primeira justificativa parte da ideia de que, independentemente da direção a que a técnica moderna se encaminhe, ela não conduz a um ponto de equilíbrio objetivando adequar os meios aos objetivos pré-fixados; ao contrário, o êxito conduz a outros passos para todas as direções possíveis. Neste sentido vale destacar que as direções possíveis incluem aquelas não previstas e não desejadas.

A segunda justificativa se refere à rapidez e à segurança com a qual a inovação técnica e as teorias científicas se difundem, e com um tempo escasso para a apropriação tanto prática como teórica de cada inovação técnica propulsionada pela concorrência no contexto do mercado capitalista.

A terceira justificativa se refere à relação entre os fins e os meios (no âmbito da ação humana) que até então se apresentava de maneira linear (do objetivo a ser alcançado ao meio que permitisse a realização do objetivo) e que passa a apresentar-se sob outras formas, em que frequentemente o uso possível da tecnologia ou seus vários fins não são antevistos por seu idealizador. Ressalta Jonas (2013) que frequentemente a tecnologia moderna inspira, produz ou "força" novos objetivos em que ninguém havia pensado previamente: "ninguém desejava ver um coração aberto em uma sala cirúrgica, [...]beber café em um copo de papel descartável ou [...] ver agentes clonados transitando entre nós" (Jonas, 2013, p. 30).

Na quarta justificativa, o autor enfatiza a mudança de significado do conceito de progresso que na tecnologia contemporânea é algo além de uma opção, ele é "impulso incerto [...] que repercute no automatismo. [...] Progresso não é, nesse sentido, um conceito valorativo, mas puramente descritivo. Podemos lamentar seus feitos e detestar seus frutos e mesmo assim temos que avançar com ele [...]" (Jonas, 2013, p. 31). Além disso, o conceito de *progresso* não pode ser entendido de modo *neutro*, no sentido de poder ser compreendido como sinônimo de "mudança", devido a cada estágio posterior ser superior ao precedente. Tais constatações visam explicar a afirmação de que "a moderna tecnologia, diferentemente da tradicional, é uma empresa e não uma posse, um processo e não um estado, um impulso dinâmico e não um arsenal de ferramentas e habilidades." (Jonas, 2013, p.32).

Jonas (2013) trata dos impulsos e das coações do próprio progresso tecnológico. Ao se referir aos impulsos, o autor aponta a pressão pela concorrência seja pelo benefício, poder, prestígio, entre outras motivações possíveis dos envolvidos na criação e implementação das novas tecnologias. Além disso, o fator econômico e os interesses do mercado sobre a nova técnica a ser desenvolvida também têm um valor relevante (embora tal pressão possa ocorrer em um sistema econômico socialista, segundo Jonas (2013)). Além da concorrência, o aumento da população e a ameaça de esgotamento de reservas atuam também como uma forma de pressão para o desenvolvimento de novas tecnologias: as pesquisas em torno aos organismos transgênicos ilustra esse tipo de pressão. Entretanto, o autor enfatiza que frequentemente "a própria técnica cria problemas que depois tem de resolver mediante

um novo salto adiante. (A "revolução verde" e o desenvolvimento de sucedâneos sintéticos ou fontes de energia alternativas são exemplo disso.)". (Jonas, 2013, p. 33).

Desse modo, segundo Jonas, o impulso para o progresso tecnológico tem se mostrado na modernidade de maneira autônoma, promovendo uma visão utópica de uma vida supostamente melhor, na qual a tecnologia contemporânea aparenta ter a capacidade de criar condições para isso de maneira contínua. Outro aspecto ressaltado por Sganzerla (2011) ao comentar esta tese jonasiana é sobre o "*encantamento moderno* com os benefícios em termos de bens e ambientes de liberdade [...]" (p.116, grifo nosso). Dessa forma, atividades que pareciam impossíveis e improváveis tornam-se realidade, mas muitas delas têm como consequência vários resultados indesejáveis como a progressiva exaustão dos recursos naturais e a transformação dos seres vivos em geral em mercadoria. Nesse viés, entre todas as possibilidades causais do progresso tecnológico contemporâneo, certamente a premissa de que "pode haver um progresso ilimitado, porque sempre há algo novo e melhor para ser encontrado" tem sido um dos discursos mais fortes para que a busca pelo progresso continue. (Jonas, 2013, p. 35). Mas Jonas chama a atenção para a diferença entre um potencial contínuo de progresso sucessivo (expresso na lógica da criação de novas tecnologias para satisfazer necessidades resultantes das tecnologias anteriores) e a efetiva melhoria das condições de vida a longo prazo por meio do desenvolvimento de novas tecnologias.

No que tange as características do progresso contemporâneo, Jonas (2013) afirma que se destaca a inter-relação entre a ciência e a tecnologia. "Se a arte tecnológica segue os passos da ciência natural, adquirirá também desta fonte aquele potencial da infinitude para suas progressivas inovações". (Hans Jonas, 2013, p. 37). Cabe ressaltar que, para alcançar seus objetivos, a ciência tende a necessitar de tecnologias cada vez mais refinadas. O autor enfatiza que ciência e tecnologia estão infiltradas uma na outra, de maneira que cada uma necessita de e impulsiona a outra.Na próxima seção aprofundaremos a noção de progresso tecnológico.

2.2- O conteúdo substancial da tecnologia segundo Hans Jonas

No que tange ao conteúdo substancial da tecnologia, Hans Jonas (2013) introduz duas observações. A primeira consiste em que a busca pelo *saber*, até então dividida em teoria e prática, não se aplica à produção da tecnologia contemporânea, pois

a "[...] autossuficiência da busca pela verdade por si mesma desapareceu. (Jonas, 2013, p. 39). Além disso, a contemplação perdeu espaço na ciência para a exploração de trabalho ativo na ciência. Enfim o progresso do ser humano acaba resumindo-se, segundo Jonas, a ampliar a busca por poder sobre a natureza independentemente de considerar implicações de tal busca a longo prazo, especialmente no campo ético.

Tal progresso é apresentado por Hans Jonas (2013) em uma perspectiva histórica, subdividida em cinco estágios: mecânico, químico, elétrico, eletrônico e biológico. Além destes, podemos acrescentar um sexto estágio, o da Inteligência Artificial capaz de produzir sistemas autônomos em algum grau não trivial.

O primeiro estágio é denominado **estágio mecânico** e é caracterizado por ser o primeiro passo para o desenvolvimento tecnológico. No contexto histórico, estágio se apresenta fortemente na Revolução Industrial no século XVIII. Nessa fase os objetos da técnica moderna eram os mesmos objetos da habilidade e do trabalho humano, de modo que não mudava o produto, mas a produção era mais rápida e fácil. A partir das novas indústrias abriram-se as indústrias auxiliares para assessorar as condições para o funcionamento da nova tecnologia. Logo os produtos finais que chegavam ao consumidor deixaram de ser os mesmos embora cumprissem o mesmo papel. A transformação dos produtos industrializados se deu de tal modo que as matérias primas não são mais reconhecidas quando transformadas no produto processado.

O **estágio químico** se caracteriza pela possibilidade de "Inferir, alterar e redesenhar os próprios padrões naturais, gerando um novo âmbito de artificialidade" (Jonas, 2013, p.14) na produção de alimentos, fibras para a confecção de tecidos, as mais variadas substâncias empregadas em diversos ramos da indústria, entre outras.

O **estágio elétrico** é iniciado pela implementação dos sistemas de distribuição de energia elétrica graças aos quais as máquinas[19] se converteram em artigos de uso pessoal (geladeiras, fogões, televisão, dentre centenas de outros produtos), utilizadas para realizar tarefas e para entretenimento. Entretanto o autor apresenta aparatos que não "fazem nenhum trabalho (físico) para nós", como é o caso dos telefones, rádios, gravadores, entre outros. Segundo o autor, a partir da difusão dessas máquinas, surgem técnicas da transmissão de informação que podem ser equiparadas a uma segunda

[19] A definição de máquina é "no sentido exato de fazerem um trabalho transformando energia em movimento mecânico e por suas partes móveis pertencem à magnitude familiar de nosso mundo sensorial" (Hans Jonas, 2013, p. 45).

revolução tecnológica. A eletricidade foi utilizada primeiramente através da telegrafia, mas posteriormente foi mais explorada e constatada que "a natureza dessa nova energia era em si mesma revolucionária. Sua distinção consistia em sua mobilidade única, a facilidade de sua transmissão, transformação e distribuição". (Jonas, 2013, p. 47).

O **estágio eletrônico** "[...] descarta definitivamente a ideia de uma imitação da natureza, para inventar objetos, objetivos e necessidades próprias" (Jonas, 2013, p. 14). A eletrônica se apresenta como uma nova revolução científica - tanto na perspectiva prática como teórica, especialmente pelas novas possibilidades que as tecnologias processadoras de informação oferecem no campo da indústria, da comunicação, do entretenimento, dentre inúmeros outros.

E o **estágio biológico**, cujas tecnologias propiciam a busca pelo controle direto sobre os processos evolucionários e as estruturas constituintes da natureza humana. Embora motivado pela superação das doenças e a busca pela saúde, as biotecnologias também propiciam, segundo Jonas (2013), utilizações eticamente polêmicas. Exemplos atuais do uso questionável da biotecnologia são a clonagem de seres humanos e a manipulação genética com fins puramente estéticos.

Por fim, acreditamos haver um **sexto estágio** constituído pelo **desenvolvimento de sistemas artificiais** dotados de algum grau de autonomia e por novas tecnologias informacionais que pretendem, inclusive, "melhorar" a condição humana por meio de interfases do ser humano/máquina (Bostrom, 2012). Não é nosso objetivo aqui analisar a fundo as teses do projeto transhumanista proposto por Bostrom (2012), entre outros, cabendo apenas ressaltar que se trata de projeto polêmico, até porque a própria noção do que efetivamente constituiria um melhoramento da natureza humana é ambígua e imprecisa.

No que se refere aos sistemas artificiais autônomos desenvolvidos na atualidade, os quais serão tratados mais detalhadamente no Capítulo 3 desta dissertação, cabe agora caracterizá-los brevemente como sendo aqueles que conseguem aprender com a experiência passada e realizar tarefas não previstas em sua programação inicial, pelo menos não inteiramente.

Tendo apresentado brevemente as teses de Jonas sobre o desenvolvimento da tecnologia, acrescentando um sexto estágio sobre o desenvolvimento dos sistemas artificiais dotados de algum grau de autonomia, passamos agora a discutir aspectos em

torno à noção clássica de responsabilidade e as teses sobre o princípio da responsabilidade propostas por Hans Jonas (2006).

2.3 A ação humana responsável

Iniciaremos esta seção apresentando alguns conceitos sobre a noção de responsabilidade. Segundo Andrew Eshleman (2014), para responsabilizar moralmente uma pessoa por um evento particular é necessário que a mesma tenha exercido algum tipo de influência sobre esse evento, se o agente não puder agir de forma diferente ou evitá-lo, não faz sentido responsabilizá-lo. Aqui pressupõe que, para se ter uma ação responsável, é necessário ter uma ação autônoma.

Nessa perspectiva, tradicionalmente a responsabilidade moral tem como foco os agentes humanos e se refere às ações humanas e suas consequências diretas. Noorman (2014) atenta que atualmente há dificuldade de atribuir responsabilidade devido ao fato dos artefatos tecnológicos terem uma interferência direta nas ações e decisões humanas. Tal interferência dificulta identificar quem deveria ser culpado ou compensado por um determinado resultado, como apontamos esta dúvida ocorre quando se trata de sistemas sociotécnicos, no sentido proposto por Smiley (2011), a que nos referimos no Capítulo 1. Os computadores e a rede de internet, em especial, transformaram a forma de interação e de comunicação entre os seres humanos.

Uma das maiores áreas de aplicação da computação, segundo Noorman (2014) é a automação dos processos da tomada de decisão e controle. A automação pode ajudar a centralizar e aumentar o controle sobre vários processos, enquanto que limita o poder discricionário dos operadores humanos na extremidade inferior da cadeia de tomada de decisão. Um exemplo é o bloqueio anti-álcool presente em alguns automóveis que já estão em uso em uma série de países, incluindo os EUA, Canadá, Suécia e Reino Unido. Tal bloqueio consiste em um sistema que exige que o motorista passe por um teste de respiração antes que possa ligar o carro. Mesmo que tenha em vista a diminuição da quantidade de acidentes provocados por motoristas embriagados, essa tecnologia obriga um determinado tipo de ação e deixa o motorista com quase nenhuma possibilidade de escolha, embora seja possível fraudar o teste pedindo para outro indivíduo assoprar no sensor em seu lugar.

A liberdade para agir é uma importante condição para a atribuição de responsabilidade moral a uma ação intencional no sentido não trivial sugerido por Frankfurt (1978). No entanto, há pouco consenso sobre quais capacidades os seres humanos têm que lhes permitem agir livremente. Como vimos, várias concepções de ação consideram que para uma ação poder ser considerada intencional e autônoma, o agente deve possuir características como a racionalidade e capacidades linguísticas.

Independentemente das discussões conceituais sobre a natureza da ação intencional, do ponto de vista da organização social humana, cabe ressaltar a diferença básica suposta entre a responsabilidade no âmbito moral e a responsabilidade jurídica. Silva (2013) aponta uma diferença inicial entre ambas: enquanto a responsabilidade moral corresponde a um âmbito de interesse individual, em que o agente deve se sentir responsável perante sua própria consciência, a responsabilidade jurídica ocorre quando há uma infração na norma jurídica civil ou penal, perturbação na paz social, algum dano a um indivíduo, uma coletividade, ou ambos. Esse indivíduo causador do dano ou do crime deve ressarcir o dano causado através de indenizações ou penas, para que essa ação que desequilibra a paz não volte a se repetir e sirva de exemplo para que ninguém mais o imite.

A responsabilidade moral se distingue então da responsabilidade jurídica no aspecto de que na primeira não há coerção e o Estado não pode exigir seu cumprimento. Reis (2014) nos apresenta o seguinte exemplo: um católico fervoroso, ao cometer um pecado e se arrepender de tê-lo cometido, confessa-o ao padre e recebe suas "punições" de caráter psicológico, como por exemplo, rezar 10 Pai Nossos. Já no caso de ações que firam a ordem social, direitos de outrem ou a integridade moral e física de outra pessoa geram responsabilidade jurídica do agente. O autor nos apresenta uma definição de responsabilidade no direito.

> Responsabilidade, para o Direito, nada mais é, portanto, que uma obrigação derivada – um dever jurídico sucessivo – de assumir as consequências jurídicas de um fato, consequências essas que podem variar (reparação dos danos e/ou punição pessoal do agente lesionado) de acordo com os interesses lesados. (REIS, 2014, p.1).

Assim, acreditamos que por enquanto nenhuma dessas concepções de responsabilidade podem ser atribuídas aos agentes artificiais, pois ainda não é possível afirmar que um agente artificial tenha consciência, mesmo no sentido mais básico da

palavra. Além disso, punir um agente artificial de modo análogo às punições aplicadas a seres humanos parece não ser muito coerente nem funcional. Notamos que os conceitos de responsabilidade moral e jurídica só puderam ser introduzidos depois que os seres humanos se organizaram em sociedade, procurando estabelecer normas éticas e jurídicas de boa convivência mútua. Mas atualmente, as interações entre seres humanos, especialmente nas sociedades industrializadas, são encontram frequentemente intermediadas pela tecnologia. Será que essa mudança traz implicações éticas significativas? Será que é necessária uma nova reflexão ética? Acreditando em uma resposta afirmativa a essas questões discutiremos na próxima parte o *princípio responsabilidade* proposto por Hans Jonas.

2.4- O conceito de responsabilidade proposto por Hans Jonas

Tendo em vista o nosso objetivo de analisar a ética da reponsabilidade proposta por Hans Jonas (2006), neste primeiro momento é importante ressaltar alguns pontos importantes da vida do autor, pois o período em que ele viveu está profundamente relacionado à sua teoria ética, que analisaremos neste capítulo.

De acordo com Battestin e Ghigg(2010), o pensamento de Hans Jonas (1903-1993) foi impulsionado por sua origem judia e pelo período em que viveu e presenciou a crise europeia nas décadas de 1920 e 1930, a Primeira e a Segunda Guerras Mundiais e o início da constituição da sociedade tecnológica no sentido contemporâneo. A partir de uma análise reflexiva de toda essa realidade de acontecimentos e destruições até então inimagináveis, Jonas (2006) concebe o conceito de responsabilidade na tentativa de promover uma visão ética que busque evitar a ocorrência de conflitos e guerras ainda piores pudessem ocorrer. Tais guerras e conflitos, promotores e usuários de tecnologias bélicas cada vez mais sofisticadas e poderosas, poderiam ser capazes de desencadear a destruição do ser humano e da natureza tal como a conhecemos. Por essas razões é necessário retraçar a problemática proposta pelo autor a respeito dos desenvolvimentos tecnológicos.

Nesse sentido, Jonas (2006) traz subsídios para podermos refletir sobre possíveis implicações a longo prazo de tecnologias contemporâneas. Jonas acompanhou a destruição causada pelas guerras mundiais e, a partir dessa experiência, propõe uma reflexão sobre o aprimoramento das tecnologias. Já no primeiro capítulo da obra

Princípio responsabilidade, o autor analisa a relação humano – poder – natureza. Sob essa perspectiva, Jonas (2006) propõe que o poder do ser humano é intensificado com o desenvolvimento tecnológico, quando se refere à intervenção humana em processos naturais. É a partir dessa reflexão que ele propõe a necessidade de se desenvolver uma nova teoria ética: as ações humanas e as relações sociais mudaram como resultado da mediação tecnológica. Podemos observar, por exemplo, o corte de árvores. Antes realizado com machado, demandava mais tempo e mais esforço físico humano se comparado ao corte de árvores realizado atualmente com motosserras. Tal aprimoramento tecnológico, por demandar menos esforço físico, potencializa o corte massivo de árvores para uso humano, o que contribui para acirrar o efeito estufa. Ademais, cabe ressaltar que a motosserra necessita de combustível, cujo uso colabora ainda mais com a poluição do ar.

A fim de analisar a relação do ser humano com a tecnologia no decorrer da história ocidental, Jonas (2006, p. 31-32) recorre ao texto o *Coral de Antígona* de Sófocles, no qual ressalta e analisa um discurso sobre o poder e o fazer humano no mundo clássico. Jonas (2006) destaca nesse texto a supervalorização do ser humano, "acima do mar, dos ventos", além da capacidade humana de capturar e domesticar os animais não humanos. No trecho analisado por Hans Jonas (2006) é nítida a noção de que os seres humanos só são impotentes contra a morte, embora tenham desenvolvido medicamentos para retardá-la.

Outro aspecto que Jonas (2006, p.33-34) observa na diferenciação entre a relação com a tecnologia do "ser humano antigo" para com o "ser humano moderno e contemporâneo" é a utilidade da construção das cidades. Segundo o filósofo, as cidades tinham como objetivo cercar os seres humanos e protegê-los do que pudesse ameaçá-los, sejam inimigos ou adversidades naturais. Nesse contexto, a natureza permanecia e as obras humanas mudavam, no entanto, nenhuma mudança era para durar, pois o controle do ser humano sobre as forças naturais era pequeno e a natureza se impunha.

Nessa perspectiva, Jonas (2006) ressalta que as ações humanas se concentravam nas cidades, sem alterar tão radicalmente os processos naturais fora delas. Atualmente, podemos observar sem dificuldade as profundas alterações nas paisagens naturais decorrentes das interferências das ações humanas. Ao observar, por exemplo, a situação atual do rio Tietê na cidade de São Paulo, que deveria ser o *habitat* de várias

espécies aquáticas e que deveria fornecer água aos animais que vivessem em seu entorno, constatamos o impacto da poluição gerada pelos habitantes da cidade.

Jonas (2006) argumenta que quando a interferência humana não afetava a natureza de modo "tão brusco", ela era resiliente o bastante para recuperar certo equilíbrio. Simultaneamente à aglomeração humana nos espaços urbanos, Jonas ressalta que foi preciso algum modo de organização moral desse novo modo de vida. "[...] na cidade onde seres humanos lidam com seres humanos é necessário moralidade, pois essa é a alma de sua essência" (Hans Jonas, 2006, p. 34). De modo que se tornou uma necessidade refletir moralmente as ações dos seres humanos entre si e perante a natureza, justamente pelo "grau de poder" desenvolvido pelo ser humano com o auxílio das tecnologias de interferir e/ou modificar a natureza em um curto período de tempo.

Enquanto o conhecimento e o poder dos humanos sobre a natureza eram limitados, de acordo com Jonas (2006) não havia necessidade de se pensar em uma ética preocupada com o futuro longínquo. A função da ética até então seria apenas se preocupar com os direitos de seus coetâneos. Com o desenvolvimento tecnológico moderno e contemporâneo e com o novo "poder" sobre a natureza, foi necessário que a ética proposta até então fosse repensada e sua abrangência ampliada e que a mesma, segundo Hans Jonas (2006) deveria adquirir uma nova necessidade e um novo dever: *preocupar-se com os seres humanos em uma projeção causal para o futuro* e com o impacto a longo prazo das opções tecnológicas contemporâneas. Assim, a noção de responsabilidade acaba por se tornar o centro da reflexão ética para a sociedade moderna e contemporânea segundo Jonas.

Deste modo, Jonas (2006, p.31-34) conclui que a violação da natureza e o florescimento da civilização caminham juntos. Será que essa afirmação procede? Na contemporaneidade, a sociedade da informação alcançou um alto grau de sofisticação tecnológica, proporcional à intervenção cada vez mais ampla e profunda nos processos naturais.

Além do ser humano ter essa forte interferência direta na natureza externa com o auxílio das tecnologias, ele também tem desenvolvido tecnologias que os afetam diretamente que são sempre desenvolvidos com o objetivo de "melhorar" e /ou "prolongar" a vida. Jonas (2006) problematiza a finalidade desses tipos de tecnologias.

Nos estudos desenvolvidos para o prolongamento da vida, o envelhecimento é considerado uma doença e não um processo natural do ciclo vital de todas as espécies. Tecnologias desenvolvidas no contexto do transhumanismo, por exemplo, já ambicionam a imortalidade, o que levanta uma grande quantidade de problemas éticos. Quem terá o direito à imortalidade? Ricos? Inteligentes? Fortes? A resposta mais adequada seria que todos os que a desejassem tivessem o direito à imortalidade. Todavia, se todos os seres humanos vivos em certa época conquistassem a imortalidade, isso impediria o nascimento de novos seres humanos. Nesse sentido, cabe questionar se poderíamos tirar o direito da existência de novos seres humanos.

Outro tema problematizado por Jonas (2006) diz respeito às manipulações genéticas. Seria o ser humano capaz de criar seres humanos? Qual modelo seria utilizado? Além de questionar a capacidade do ser humano para tal Jonas (2006) questiona também se criar, modificar ou "melhorar" seres humanos é uma atitude correta.

Nessa perspectiva Jonas (2006) também problematiza a possibilidade de controle do comportamento por meio de drogas, por exemplo. Se por um lado pode parecer positivo utilizá-lo como instrumento para aliviar a dor e as perturbações de doentes mentais, por exemplo, por outro lado, as empresas poderiam se apropriar dessa técnica para aprimorar a performance e manipular seus funcionários, com estímulos cerebrais de prazer, por exemplo. Além disso, seria positivo tirar todos os "problemas" e "dores" da vida dos seres humanos? Lembremos que quando o corpo manifesta dor, ele sinaliza que há algo de errado com a saúde da pessoa.

Além do mais, ao contrastar processos naturais e o desenvolvimento da técnica, Jonas (2006) ressalta que a probabilidade de sucesso de intervenções humanas mediadas por tecnologias em tais processos é muito pequena. A evolução trabalha com pequenos passos, pequenas apostas e lentamente de modo a resultar em pequenos erros e pequenos acertos. A tecnologia moderna, porém, anda rapidamente, em passos colossais e "despreza a vantagem daquela marcha lenta da natureza, cujo tatear é uma segurança para a vida". (Jonas, 2006, p. 77), trazendo ao ser humano novos perigos, que muitas vezes passam despercebidos.

Um aspecto que podemos notar, por exemplo, são as imagens dos grandes centrosurbanos onde fica muito visível a pressa com que as pessoas se locomovem, a

pressa das crianças em se vestirem e agirem como adultos, a pressa em conseguir sucesso profissional e acadêmico, a pressa para fazer uma refeição, o ser humano, de maneira geral, parece sempre estar com pressa no mundo contemporâneo, parte dela aparentemente imposta pelo ritmo das tecnologias contemporâneas.

Após esta breve introdução sobre algumas questões que Jonas (2006) aborda sobre o aprimoramento das tecnologias, e as novas preocupações advindas desses aprimoramentos, analisaremos mais adiante o conceito de responsabilidade juntamente com a proposta ética jonasiana.

Jonas (2006) aponta que o exercício de uma ação carregada de propósito e efetuada sem coação tem a responsabilidade sobre suas consequências como contrapartida. Desse modo o autor apresenta a relação dos conceitos de liberdade e responsabilidade do ser humano para com a natureza visto que com o poder humano de destruição alcançado na modernidade e contemporaneidade põe em risco toda a natureza e faz com que nós sejamos os principais responsáveis pela mesma

> O privilégio da liberdade carrega em seus ombros o fardo da necessidade, e significa existência em risco. Pois a condição básica para o privilégio consiste no fato paradoxal de a substância viva, por um ato primordial de isolamento, se haver desprendido da integração geral das coisas no todo da natureza, de haver-se oposto ao mundo, com isso introduzindo na segurança indiferente da posse da existência a tensão entre o ser e o não ser" (Jonas, 2004, p.14).

Além dessa caracterização, Jonas (2006) especifica mais o conceito de responsabilidade de que ele fará uso e apresenta algumas distinções. Na medida em que o agente tem a capacidade de iniciar uma cadeia causal no mundo, ele é responsável pelas implicações de suas ações. "O poder causal é a condição da responsabilidade" (Jonas, 2006, p. 165). O agente deve responder por suas ações, independentemente se foram causados por ações moralmente contestáveis ou por ações cujas consequências não tenham sido previstas e/ou desejadas (basta o agente ter sido a causa ativa). "Mas isso somente se houver um nexo causal estreito com a ação, de maneira que a imputação seja evidente e suas consequências não se percam no imprevisível." (Jonas, 2006, p.165). O autor apresenta o exemplo do aprendiz de ferreiro que se esqueceu de colocar um prego na ferradura de um cavalo que ia ser usado em combate. O aprendiz não poderia ser culpado pela derrota da batalha e a consequente perda do reino. Todavia, o

soldado poderia queixar-se com o ferreiro (o "responsável") que é a pessoa quem designou a um aprendiz a tarefa de pôr a ferradura em seu cavalo.

Jonas (2006) parte do princípio de que o poder causal torna o agente responsável pelas ações de um subordinado, por exemplo, assim como os pais são responsáveis por seu filho, e se responsabilizam por suas ações respondendo a alguma acusação ou ganhando mérito pelo bom desempenho.

Hans Jonas (2006, p.165-174) também considera a distinção básica entre responsabilidade moral e legal e analisa todo um conjunto de outras distinções como, por exemplo, a responsabilidade como uma ação não recíproca; a responsabilidade natural e a responsabilidade contratual e contrasta a responsabilidade do homem político com a responsabilidade parental.

A análise jonasiana de responsabilidade ressalta o papel direcionados da ação moral que ela assume. O objetivo é zelar pelo objeto externo que depende do agente, por necessidade ou que é ameaçado pelo agente. O agente deve responsabilizar-se por aquilo que está em seu poder. "Ao meu poder ele, (o objeto que está sob minha responsabilidade) contrapõe o seu direito de existir como é ou poderia ser, e com a vontade moral ele submete o meu poder" (Jonas, 2006, p. 167). O sentimento de responsabilidade é consequência da relação causal daquele que está no dever ser do objeto, para com o responsável pelo objeto.

> A reivindicação do objeto, de um lado, na insegurança de sua existência, e a consciência do poder, de outro, culpada da sua causalidade, unem-se no sentimento de responsabilidade afirmativa do eu ativo, que se encontra sempre intervindo no Ser das coisas (Jonas, 2006, p.167).

Esse sentimento de responsabilidade pode se unir ao sentimento de amor e resultar naquilo que o autor denomina responsabilidade não-formal e vazia. Jonas (2006, p.68) afirma que os dois termos de responsabilidade tendem a assegurar a afirmação de que alguém é responsável até por seus atos irresponsáveis. Mas o que é agir de forma irresponsável? Podemos antecipar que agir irresponsavelmente consiste em agir sem antever e refletir sobre as consequências futuras da ação presente ou sem considerar tais consequências. Desse modo, um idealizador de alguma recente tecnologia seria irresponsável se divulgasse seu projeto sem considerar as consequências que a proliferação de tal tecnologia poderia causara longo prazo.

A responsabilidade segundo Hans Jonas (2006, p. 168-169), pode ser assumida de modo duradouro, como no caso dos pais que assumem a responsabilidade de proporcionar condições básicas de sobrevivência a seus filhos, mas pode ser também uma responsabilidade temporária, limitada, como a responsabilidade de um motorista de ônibus que se compromete a zelar pelo bem-estar de seus passageiros durante uma viagem. Neste sentido, seria problemático se os criadores de novas tecnologias alegassem possuir apenas uma responsabilidade temporária, argumentando que têm pouco poder sobre as possíveis aplicações, especialmente no campo bélico, de suas criações. Mas também parece igualmente problemático atribuir-lhes responsabilidade duradoura, especialmente quando se trata de componentes cujos possíveis usos não sejam facilmente previsíveis. Considerando o exemplo anterior, a possibilidade de responsabilização dos pais pelas ações de seus filhos, mesmo em se tratando de uma responsabilidade duradoura, termina na maioria dos países ocidentais quando estes atingem a maioridade.

Jonas (2006) afirma que "o exercício do poder sem a observação do dever é, então, irresponsável, ou seja, representa uma quebra de confiança presente na responsabilidade" (Jonas, 2006, p.168). Por exemplo, um agente responsável pelo sustento de uma família, pode ser responsabilizado quando decide fazer uma aposta (independentemente de ganhar ou perder) ao invés de fazer uma compra de alimentos para sua família. O mesmo ocorre quando um motorista de ônibus põe a vida de seus passageiros em risco ao dirigir com imprudência e quando o criador de novas tecnologias desenvolve sistemas artificiais bélicos, por exemplo, que sabidamente ameaçam vidas humanas. O dilema moral emerge, porém, quando tais sistemas, embora ameacem por princípio vidas humanas, são desenvolvidos e utilizados para realizar tarefas que, no campo de guerra, minimizam perdas humanas de um dos lados do conflito e, alegadamente, ao facilitar a vitória dos seus desenvolvedores, acabam por poupar vidas de ambos os lados. Este argumento, por exemplo, foi utilizado para justificar o lançamento de bombas atómicas sobre as cidades japonesas de Hiroshima e Nagasaki.

A responsabilidade natural e a responsabilidade contratual, por sua vez, se diferenciam segundo Jonas (2006, p.170-171) pelo aspecto de que a primeira não depende da concordância prévia do agente, ao contrário da segunda. Jonas ressalta que a

responsabilidade parental é exemplo da responsabilidade que emerge naturalmente. A responsabilidade contratual ou artificial "... é instituída a partir da atribuição e aceitação de um encargo, por exemplo, a assunção de uma função (mas também aquela resultante de um acordo tácito ou da competência) [...]" (p.170). Exemplo deste tipo de responsabilidade é a do homem público que assume um cargo político. A responsabilidade decorrente de se candidatar a um cargo público (político) é livremente escolhida em estados democráticos, ninguém é obrigado a se candidatar. No entanto, quem o faz e vence a disputa, o faz com a pretensão de obter um poder social mais abrangente e consequentemente assumir uma responsabilidade mais ampla pelos seus atos. Jonas (2006, p.172) nota que essa responsabilidade não deve ter como essência o "sobre", mas sim, a responsabilidade "para" com aqueles em relação aos quais o governante detém o poder. Embora a responsabilidade natural e a responsabilidade do político se situem em extremos opostos, há características comuns entre elas, sendo, segundo Jonas (2006, p.173-174) as que mais têm características em comum e que mais contribuem para a compreensão da natureza da responsabilidade. Para dissertar sobre essa semelhança, Jonas (2006, p.175-187) dedica uma seção denominada *Teoria da responsabilidade: "pais homens de Estado como paradigmas eminentes*", cujos elementos centrais discutiremos a seguir.

O autor enumera três conceitos que resumem as características comuns à responsabilidade parental e à responsabilidade política: "totalidade, continuidade e futuro" (Jonas, 2006, p.175). Para poder responsabilizar o agente, Jonas aponta que, dentre os animais, o ser humano é o único capaz de tornar-se por si só objeto de responsabilidade além de compartilhar com o outro a responsabilidade da comunidade humana.

> Todo o ser vivente é seu próprio fim, e não tem necessidade de outra justificativa qualquer. Desse ponto de vista, o homem não tem nenhuma outra vantagem em relação aos outros seres viventes, exceto a que só ele também pode assumir a responsabilidade de garantir os fins próprios aos demais seres (Jonas, 2006, p.175).

Muito embora Jonas (2006, p.175) afirme que o arquétipo da noção de responsabilidade está embasado na relação humano-humano e que há uma relação de reciprocidade entre os seres humanos e, ainda, que todos os seres viventes sejam passíveis de se tornarem objeto de responsabilidade, somente o ser humano é capaz de

ser responsável, pois é o único com faculdades suficientes para tal condição[20]. Jonas (2006) afirma com muita convicção que a existência do ser humano é inseparável da responsabilidade, assim como "lhe é inalienável a sua natureza falante" (Jonas, 2006, p.176). Logo, o autor observa que se torna um dever do ser humano ser responsável por aquilo a que está relacionado casualmente por intermédio de suas ações.

Ao ressaltar as semelhanças da responsabilidade parental e a do homem público, Jonas (2006) aponta a primeira semelhança que é de totalidade, "com isso queremos dizer que a responsabilidade abarca o ser total do objeto, todos os seus aspectos, desde a sua existência bruta até os seus interesses mais elevados" (Jonas, 2006, p.180). Na responsabilidade parental é muito visível que a questão da totalidade se trata de cuidar da criança como um todo, no primeiro momento parece que a preocupação só advém do cuidado para com o físico da criança, enquanto que um tempo depois é necessário zelar pela educação, comportamento, caráter, conhecimento (se possível a felicidade) durante toda a vida da criança.

Segundo Jonas (2006), a responsabilidade parental e a responsabilidade política se entrelaçam. Na educação da criança, os pais educam seus filhos para que possam se tornar membros da vida social e serem cidadãos. Todavia, o Estado também tem que participar da educação dos cidadãos oferecendo a educação obrigatória (atual educação básica) e interferindo quando necessário na educação familiar como, por exemplo, quando os pais são dependentes químicos e não tem condições básicas para cuidar e educar seus filhos.

O filósofo chama a atenção para o que ele denomina de coletivização extrema, proposta por abordagens políticas contrárias à noção de âmbito privado, na qual o Estado deteria o poder sobre a educação da criança e aboliria a responsabilidade paternal. Nesse sentido, Jonas (2006, p.182-183) faz uma analogia quanto ao sentimento paternal de pais para com os filhos e o sentimento "paternal" do Estado para com seus cidadãos.

Embora governantes não tenham motivo para desenvolver uma relação afetiva com os governados, é possível, e até necessário, que desenvolvam um sentimento de

[20] Embora Jonas critique vários aspectos das éticas antropocêntricas, esta tese jonasiana se enquadra no contexto teórico antropocêntrico, atualmente sob revisão, segundo o qual apenas os seres humanos, por possuírem linguagem, contam com as capacidades cognitivas necessárias para realizar operações mentais complexas que envolvem abstrações.

solidariedade, de maneira que o governante se sinta como filho daquela pátria e crie um sentimento de irmandade, solidariedade para com aqueles que vivem no mesmo lugar e compartilham de muitas coisas em comum. Em defesa da importância de o governante manter um vínculo de solidariedade e empatia com os governados, Hans Jonas (2006) defende que é difícil, senão impossível, assumir a responsabilidade por algo emocionalmente indiferente para o agente.

Outra analogia feita por Jonas (2006) entre os vínculos de responsabilidade parental e política é que, assim como a criança, a comunidade necessita de alguém que tome conta dos assuntos políticos que lhe dizem respeito, pois eles não se desenvolvem nem se solucionam sozinhos. Mas, ao contrário da responsabilidade parental, em que o foco é a responsabilidade sobre outrem a criança, na responsabilidade política, o governante, enquanto membro da comunidade está defendendo também seus próprios interesses e supre as próprias necessidades.

No que se refere aos criadores e desenvolvedores dos sistemas artificiais, consideramos que o idealizador teria a responsabilidade de acompanhar o seu projeto como um todo, com um sentimento "paternal", e ao mesmo tempo também considerar as possíveis implicações de suas criações para a coletividade. Deste modo, o idealizador deveria considerar as possíveis implicações futuras de suas criações, evitando propor um projeto que favoreça poucos e prejudique muitos, por exemplo.

A continuidade resultante da responsabilidade política segundo Jonas (2006) exige do político uma reflexão histórica sobre sua comunidade, devendo interrogar-se sobre a trajetória passada de sua comunidade e quais serão os possíveis rumos futuros. A responsabilidade paterna é duplicada, no sentido que além de haver de se preocupar com o futuro da criança e com toda sua historicidade, é necessário também preparar a criança para a vida em sociedade. "Com isso, o horizonte da continuidade amplia-se no mundo histórico; uma se sobrepõe ao outro, e assim é impossível à responsabilidade educativa deixar de ser "política", mesmo no mais privado dos âmbitos" (Jonas, 2006, p.184). O mesmo se aplica aos idealizadores, com tantos experimentos, sistemas e tecnologias já criados. É necessário ter um estudo histórico dos resultados obtidos tanto negativos como positivos e pensar na continuidade deste progresso tecnológico, há necessidade de rever as normas éticas e jurídicas para a implementação de novas tecnologias? Este e outros questionamentos são necessários para garantir a continuidade

de progresso tecnológico, mas sem comprometer a coletividade humana, especialmente as gerações futuras.

Tendo apresentado as características de totalidade e continuidade, tratamos agora da última delas, qual seja, a preocupação com o futuro. Para Jonas (2006), na responsabilidade para com a vida, o futuro é também, até mais do que o presente imediato, objeto de preocupação. É justamente com o futuro da vida no contexto das novas tecnologias que o autor está preocupado e revela isso em sua teoria ética.Se os pais devem se preocupar com o futuro de seus filhos e os políticos com o futuro dos cidadãos, os criadores e desenvolvedores de novas tecnologias deveriam ter que se preocupar com o impacto futuro previsível que suas invenções podem ter.

Nesse viés, Jonas (2006) propõe o princípio responsabilidade entrelaçado a sua teoria ética. De modo bem simples, podemos afirmar que Jonas (2006) utiliza-se do princípio responsabilidade como a necessidade de preocupar-se com as implicações futuras das ações presentes, isto é, com os efeitos das ações presentes para as gerações vindouras. Jonas sugere que sejamos humildes em um novo sentido "... não como a [humildade] do passado, em decorrência da pequenez, mas em decorrência da excessiva grandeza do nosso poder [...]" (Jonas, 2006, p.63).

O conhecimento e o poder dos seres humanos eram mais limitados no passado, por isso não havia necessidade de se pensar uma ética preocupada com o futuro longínquo. A função e a necessidade da ética até então eram apenas preocupar-se com os direitos de seus coetâneos. Com o desenvolvimento tecnológico moderno e contemporâneo, e seu novo "poder" sobre a natureza fez com que a ética adquira uma nova necessidade e um novo dever: preocupar-se com os seres humanos numa projeção causal para o futuro. Desse modo, a responsabilidade em relação às consequências futuras da ação presente passa a ser o centro das preocupações éticas, segundo Jonas (2006).

A tecnologia potencializa o alcance da ação humana moderna, o que exige a necessidade de repensar as teorias éticas. Na visão jonasiana, os preceitos de justiça e honradez não são em geral desprezados nas ações pessoais e cotidianas, mas no caso do fazer coletivo, tais preceitos são muitas vezes esquecidos. No âmbito da ação política "[...] ato efeito e ação não são mais os mesmos da esfera próxima" (Jonas, 2006, p. 39). Um exemplo que podemos citar da ampliação do alcance da ação humana pela

mediação tecnológica é o dos efeitos da bomba atômica lançada em Hiroshima e Nagasaki. Desde 1945 a radiação resultante da bomba atômica interfere na identidade genética daqueles que tiveram seus ancestrais atingidos pelos seus efeitos.

Para uma melhor compreensão da ética tradicional que Hans Jonas critica e considera inadequada para refletir sobre o contexto contemporâneo, retomamos na sequência algumas das principais características de tais concepções éticas criticadas por Jonas (2006):

1- A tecnologia era considerada eticamente neutra, pois poderia ser utilizada para distintos fins e seu impacto sobre a natureza era bastante restrito. Além disso, a *técnica* tinha como objetivo suprir necessidades e não promover o progresso. Atualmente os cultivos transgênicos são ótimos exemplos do possível impacto da ação humana na natureza multiplicado pela inovação tecnológica, mas cujas consequências a longo prazo permanecem desconhecidas. Vale frisar que a busca é, supostamente, pelo "aperfeiçoamento" do alimento.

2- As abordagens éticas tradicionais são majoritariamente antropocêntricas, pois focalizam prioritariamente a relação do ser humano com ele mesmo e com outros seres humanos. Embora esta crítica de Jonas fosse adequada no fim dos anos de 1970, quando a obra *Princípio responsabilidade* foi publicada, atualmente há várias abordagens éticas preocupadas com o meio ambiente, as relações seres humanos-não humanos, problemas associados ao uso das novas tecnologias, entre outros. Exemplos de máximas morais tais como: "Ama o teu próximo como a ti mesmo"; "Instrui teu filho no caminho da verdade"; "Nunca trate os teus semelhantes como simples meios, mas sempre como fins em si mesmo" revelam que o universo moral focalizava as interações humanas.

3- A natureza humana tem, segundo as abordagens éticas tradicionais, certa estabilidade e permanência que não é atingível pela *techne*. No entanto, o contexto contemporâneo revela que a interação dos seres humanos com as novas tecnologias está alterando traços bastante profundos e estruturais de sua natureza bem como de suas formas de interação social. Nesse tópico podemos utilizar como exemplo as redes sociais. Qualquer um pode ser "perfeito" pela internet, pode apresentar um perfil, ou vários perfis que nem

sempre correspondem com o que a pessoa é efetivamente. Utilizando o *software* Photoshop é possível alterar a aparência física de fotografias, retirando as supostas imperfeições e ter uma aparência considerada ideal. O agente pode também expressar sua opinião sem assumir sua identidade e nem enfrentar eventuais implicações decorrentes delas, a não ser, eventualmente, em caso de crimes.

4- Nas éticas tradicionais, o bem e o mal não necessitam de uma reflexão a longo prazo e ampla, considerando aspectos temporais e espaciais. As consequências a longo prazo ficavam a cargo do destino e da sorte. As teorias éticas tradicionais se preocupavam mais com as ações humanas aqui e agora, por assim dizer. Atualmente, porém, especialmente no que tange ao impacto das novas tecnologias informacionais, bélicas e a biotecnologia, o possível impacto ético das ações humanas extrapola o tempo e o espaço próximos, com consequências imprevisíveis.

A intervenção humana na natureza fortalecida pela tecnologia contemporânea provocou uma vulnerabilidade da natureza até então inimaginável. Jonas considera que, se detemos o poder sobre a natureza, devemos ser responsáveis pelas implicações de nossas ações sobre ela, sendo preciso uma nova ética que questione que tipos de deveres serão exigidos dos agentes que partilham de tal poder.

Jonas (2006) atenta que se o interesse na manutenção da natureza se der pela dependência que o ser humano tem dela, a ética ainda estará embasada no antropocentrismo, porém com algumas diferenças, pois as éticas tradicionais não contavam com um comportamento cumulativo. Poderíamos reivindicar um estatuto moral para a natureza? Parece que é exatamente isso que Jonas (2006) está propondo.

Nas épocas que antecederam a modernidade, a técnica florescia para satisfazer necessidades básicas, não sendo considerada um fim em si mesma, apenas um meio para alcançar fins próximos, bastante definidos [...] “somos tentados a crer que a vocação dos homens se encontra no contínuo progresso desse empreendimento, superando-se sempre a si mesmos, rumo a feitos cada vez maiores” (Jonas, 2006, p. 43). Nesse aspecto devemos refletir se há limites para o progresso, o que pode e o que não pode, até onde se pode ir. Desse modo, o *homo faber* acaba superando *o homo sapiens*.

Na ética da responsabilidade também é possível a aplicação da distinção kantiana de imperativo categórico e imperativo hipotético, porém, em relação ao futuro.

De modo que o imperativo hipotético se dá de várias formas, como por exemplo, se houver seres humanos no futuro, então caberá a eles respeitar as máximas morais. O imperativo categórico supõe simplesmente que haja seres humanos que possam refletir sobre as implicações da generalização de suas ações (Jonas, 2006, p. 94-95).

Para Jonas (2006, p. 93-95), o primeiro imperativo é haver humanidade, as demais regras se submetem a essa. O autor também discorre, por exemplo, sobre o imperativo ontológico que remete a reponsabilidade que os humanos têm pela concepção de natureza humana que transmitirão (o que exige o cumprimento do primeiro dever é necessário uma corporificação do ser humano para se refletir sobre as implicações de sua conduta).

Assim como Kant, Jonas (2006) também propõe um imperativo que deve ser considerado na tomada de decisão. Para o filósofo, tal imperativo colabora para a reflexão sobre as ações dos seres humanos na sociedade contemporânea.

> "Aja de modo a que os efeitos de sua ação sejam compatíveis com a permanência de uma autentica vida humana sobre a Terra" ou expresso negativamente "Aja de modo a que os efeitos da tua ação não sejam destrutivos para a possibilidade futura de uma tal vida" ou simplesmente: "Não ponha em perigo as condições necessárias para a conservação indefinida da humanidade sobre a Terra"; ou, em uso novamente positivo "Inclua na sua escolha presente a futura integridade do ser humano como um dos objetos do seu querer" (Hans Jonas, 2006, p. 47-48).

O autor ressalta que a partir desse novo imperativo se pode arriscar a própria vida, mas não a da humanidade. Também Jonas (2006) aponta um aspecto diferencial: o novo imperativo está voltado à conduta pública enquanto o imperativo kantiano está voltado para o indivíduo em seu critério momentâneo. "Nossa tese é que os novos tipos e limites do agir exigem uma ética de previsão e responsabilidade compatível com esses limites" [...] (Jonas, 2006, p. 57). Isto posto, cabe refletir sobre os instrumentos que permitirão a essa nova ética refletir sobre as possíveis consequências futuras da ação presente no que se refere à preservação da vida. Com esse objetivo, Jonas (2006) sugere a adoção da heurística do temor, cujos principais traços passaremos a expor na próxima seção.

2.5- O conceito de *heurística do temor*

A partir do conceito de heurística do temor, Jonas (2006) busca trazer um instrumento, em uma perspectiva prática, para a efetivação do princípio responsabilidade. O autor propõe uma consciência ética guiada pelo prognóstico pessimista, "utilizar o temor como forma de aprendizado e fazer da projeção da possibilidade da previsão negativa como condição para alterar a atitude do ser humano frente à natureza" (OLIVEIRA, 2011, p. 11). Tal heurística tem como meta refletir sobre as características de um futuro que preserve as condições necessárias para a manutenção da vida.

Jonas (2006) endossa a antiga afirmação de que é melhor ter um mundo do que nenhum, porém com o progresso tecnológico se torna uma obrigação moral criar as condições ao alcance humano para a existência de uma posteridade distante. O dever era oriundo da concepção de presença do ser humano no mundo, agora este se torna o dever, conservar este mundo físico para garantir a presença do ser humano e demais seres no planeta.

Para Jonas, torna-se dever do ser humano, no presente, zelar pela existência e talvez até pela felicidade da geração vindoura, para mais tarde eles não culparem sua geração progenitora por ações descuidadas e/ou imprudentes que lhes causaram a infelicidade. Esse é então o primeiro dever da nova ética da responsabilidade: ter uma visão a longo prazo das possíveis implicações futuras das ações presentes. Como uma ética do futuro, deve temer o que ainda não foi experimentado "o *mal* imaginado deve assumir o papel de *mal* experimentado" (Jonas, 2006, p. 72). Esse é o primeiro dever ético segundo Hans Jonas e consideramos que tal dever precisaria ser considerado em relação ao contexto tecnológico contemporâneo, pois não parece haver muitas preocupações com as consequências negativas que poderão vir futuramente do uso indiscriminado dos sistemas artificiais ditos autônomos.

O segundo dever é imaginar o *mal* como se estivesse acontecendo consigo ou a um ente próximo, pois ao imaginar o *mal* de uma pessoa desconhecida do futuro não teria o mesmo efeito. Imagine uma morte violenta a qual todos temem até mesmo por instinto, se consigo "um temor de tipo espiritual, que, como resultado de uma atitude deliberada, é nossa própria obra. A adoção dessa atitude, ou seja, a disposição para se deixar afetar pela salvação ou pela desgraça (ainda que só imaginada) das gerações

vindouras [...]" (Jonas, 2006, p.72) realiza então o segundo dever. Cabe indagar, será que se os idealizadores de armas bélicas autônomas considerassem a possibilidade de terem suas famílias atingidas por suas criações, eles ainda as implementariam?

Faz-se necessário utilizar o bem geral, que visa o melhor para todos, no que se refere à preservação da natureza e da humanidade, como princípio norteador das ações. O bem não deve ser realizado de maneira subjetiva, "o bem para mim", mas, segundo Jonas, deve ser tomado como "causa do mundo". "Podemos reconhecer um bem em si na capacidade como tal de ter finalidade, pois se sabe intuitivamente que ela infinitamente superior a toda falta de finalidade do ser" (Jonas, 2006, p.150).

O *mal* não é desejado, e é facilmente percebido quando suas implicações são evidentes, enquanto o *bem* pode passar despercebido. Uma vez que, segundo Jonas, sabemos aquilo que não queremos (como dor, sofrimento, injustiça) antes de saber especificamente o que queremos, a filosofia moral deve se encarregar de investigar o que não desejamos para nós e nossos seres próximos antes do que desejamos: "embora, portanto, a heurística do medo não seja a última palavra na procura do bem, ela é uma palavra muito útil" (Hans Jonas, 2006, p. 71).

Para aplicar a heurística do temor, é preciso primeiro refletir sobre as finalidades das ações e sobre a natureza dos fins. O fim: "é aquilo graças ao qual uma coisa existe e cuja produção ou conservação exigiu que algum processo ocorresse ou que alguma ação fosse empreendida. Ele responde à pergunta: "Para que"?" (Jonas, 2006, p. 107). Por exemplo, o fim do martelo é martelar, de um tubo digestivo, digerir, para manter o organismo vivo e em boa constituição para conseguir se locomover, a finalidade do progresso tecnológico contemporâneo seria "melhorar, progredir". Os fins definem as coisas independentemente de seu status como valor. Reconhecer um fim não significa fazer algum julgamento a respeito.

No que diz respeito à relação entre valor e fim, o autor afirma "assumo o "ponto de vista" das coisas, posso então evoluir do conhecimento de seus fins imanentes para julgamentos sobre sua maior ou menor adequação a eles, isto é, sobre a sua utilidade para a obtenção desses fins" (Jonas, 2006, p.107). Desse modo, o filósofo conclui que se pode formar o conceito de bem e de seu oposto, o mal, a partir da percepção dos fins nas próprias coisas. "É o bem conforme a medida da utilidade para um fim (cujo próprio fato de ser bom não está em julgamento)" (Jonas, 2006, p.108).

Nessa perspectiva, indagamos quais são o fim para a criação e o desenvolvimento de sistemas artificiais, ditos autônomos? Ou qual é a finalidade de atribuir autonomia e outras capacidades até então consideradas especificamente humanas a seres não biológicos? É claro que não temos respostas para tais questões, mas trataremos delas no terceiro capítulo desta dissertação.

Jonas (2006, p. 109) distingue, ainda, o duplo sentido contido na expressão "ter um fim". O martelo, como mencionado, foi criado para o fim de martelar, uma pedra pode ter um fim momentâneo de quebrar um galho para se alcançar algo, mas esses mesmos objetos podem ser utilizados para inúmeros outros fins. Nesse sentido, Jonas alerta para uma questão muito relevante: a de que os fins dos objetos não são propriamente dos objetos, mas de seus fabricantes e usuários. Dessa forma, o relógio e o martelo, assim como as tecnologias contemporâneas seriam, pelo menos em princípio, destituídos de fins em si próprios, segundo Jonas. Já os seres vivos constituiriam, segundo ele, fins em si mesmos que se trata de preservar. O progresso tecnológico ilustra bem isso: por exemplo, o avião foi construído com a finalidade de transportar pessoas, mas posteriormente ele passou a ser utilizado para fins bélicos.

Cabe ainda destacar o papel do sentimento de "temor" na concepção jonasiana de responsabilidade. Esse sentimento é um ponto primordial na ética da responsabilidade que objetiva uma visão a longo prazo [...] "nos auxilia antes de tudo à previsão de uma deformação do homem, que nos revela aquilo que queremos preservar no conceito. Precisamos da ameaça à imagem humana - e de tipos de ameaça bem determinados – para, com o temor gerado, afirmarmos uma imagem humana autêntica" (Jonas, 2006, p. 70). Desse modo, devemos, através do sentimento de temor buscar antecipar possíveis consequências a longo prazo de diferentes linhas de conduta possível.

Oliveira (2011, p. 1) afirma que o conceito de heurística do temor é um dos conceitos mais interessantes e polêmicos das obras jonasianas. A heurística do temor é colocada como uma tomada de consciência, um despertar do sentimento do medo para o perigo e o risco efetivo do mal, que advém principalmente da produção e uso desatento de tecnologias cada vez mais poderosas. Esse sentimento de temor teria a função de despertar a responsabilidade no ser humano. Oliveira (2011) chama a atenção para a

heurística do temor enquanto processo reflexivo com o objetivo de antecipar consequências nefastas para o futuro humano e planetário.

Jonas (2006) lembra que a ciência consegue fazer previsões, com maior ou menor precisão e alcance dependendo do caso, em muitas áreas de investigação, mas não tem mostrado interesse de fazer previsões sobre implicações futuras dos resultados das implementações tecnológicas de suas pesquisas, especialmente no campo ético.

Com a experiência já obtida com o desenvolvimento da tecnologia, verificou-se que ela tende a adquirir uma dinâmica própria tornando-se não só irreversível como autopropulsionada, em certo sentido, ultrapassando constantemente os fins para os quais fora inicialmente planejada. Consequentemente, alerta Jonas (2006), as correções de rumo se tornam cada vez mais difíceis, razão pela qual devemos estar atentos ao despontar dessas novas tecnologias, observar o primeiro passo, a primeira ação, pois uma vez impulsionadas pela dinâmica social é muito difícil deter sua produção, difusão, e as consequentes relação de dependência e emergência de novas necessidades. Por exemplo, será que podemos conceber o mundo contemporâneo sem energia elétrica, sistemas de refrigeração, computadores, tecnologias de uso diagnóstico, celulares, e demais tecnologias de comunicação? Parte-se do pressuposto de que essas tecnologias trouxeram benefícios para a sociedade humana que devem ser preservados, porém Jonas ressalta que é ingênuo e perigoso acreditar que o avanço tecnológico é sempre benéfico e constitui uma expressão do progresso. A prescrição prática da ética da responsabilidade afirma que "é necessário dar mais ouvidos à profecia da desgraça do que à profecia da salvação" (Jonas, 2006, p. 77).

Em síntese, segundo Hans Jonas (2006, 2013) os conceitos de autonomia e responsabilidade estão entrelaçados: agentes são responsáveis por terem a autonomia de agir levando em consideração as possíveis consequências futuras de suas ações, utilizando a heurística do temor. Em especial, a criação e implementação de novas tecnologias devem, segundo Jonas (2006, 2013) ser concomitante a uma reflexão sobre seu possível impacto a longo prazo, devendo prevalecer à prudência diante de antecipações de caráter negativo.

A questão relevante é indagar se sistemas artificiais autônomos, embora não possam ser considerados seres vivos no mesmo sentido em que os organismos o são, poderiam, devido à possibilidade de aprender com suas experiências passadas, ser

considerados agentes autônomos e, consequentemente, responsáveis. Caso efetivamente os sistemas artificiais tenham o grau de autonomia que lhes permita agir responsavelmente, no sentido aqui esboçado de terem a capacidade de atuar levando em consideração as possíveis implicações futuras de suas ações, não reconhecer sua condição de agentes parece problemático, entretanto nas reflexões éticas feitas até o momento, a responsabilidade decorrente da atividade dos sistemas artificiais ainda é atribuída aos seus criadores e implementadores, de modo análogo ao que ocorre em se tratando da responsabilidade paternal, o criador deve assumir as consequências negativas e positivas atribuídas a suas invenções. Mas, se sistemas artificiais gozam de algum grau de autonomia, como é alegado por seus criadores, não deveriam tais sistemas ser considerados corresponsáveis por possíveis implicações de suas ações? Investigaremos esta questão mais detalhadamente no próximo capítulo.

CAPÍTULO 3
Relações entre autonomia, agência e corporeidade: corpo artificial e corpo biológico

Apresentação

Este terceiro capítulo tem como objetivo analisar a possibilidade de aplicação do conceito de autonomia agentes sistemas artificias. Partindo do pressuposto de que para um agente ser moralmente responsável ele deve ter condições de escolher agir de uma maneira ou de outra autonomamente e poder deliberar a partir da reflexão sobre possíveis consequências de sua escolha, tradicionalmente a agência moral é atribuída somente a seres humanos. Uma das principais razões desta atribuição é que se considera que somente os seres humanos têm as ferramentas linguísticas, abstratas e lógicas necessárias para realizar esse tipo de reflexão, entendida como de alto nível. Entretanto, segundo Noorman (2014), na contemporaneidade o vocabulário ético se apresenta de modo limitado para pensar as dimensões morais da realidade tecnológica atual, a qual engloba tecnologias complexas que propiciam muito rapidamente novas possibilidades de ação, geradoras de novas responsabilidades.

Assim, com o objetivo de investigar a possibilidade de atribuir responsabilidade, no sentido apresentado no capítulo anterior, a sistemas autônomos artificiais, o presente capítulo está dividido em três seções. A primeira seção será uma breve problematização do conceito de autonomia e o sentido de sua atribuição a sistemas artificiais. A segunda seção analisa e discute teses centrais do texto *Robotics, philosophy and the problem of autonomy* de Willem F. G. Haselager. E a terceira seção é uma discussão sobre a relação proposta por Hans Jonas entre as noções de responsabilidade e vida e as suas possíveis implicações para a possibilidade de atribuição de responsabilidade a sistemas autônomos artificiais.

3.1- Primeiras problematizações acerca do conceito de autonomia

Embora a relação entre seres humanos e aparelhos tecnológicos seja algo corriqueiro na contemporaneidade, essa relação foi se complexificando ao longo do tempo, pois a interação entre o ser humano e a máquina é muito antiga. Susana Nascimento (2006) apresenta uma reflexão desta relação no decorrer da história que nos permitirá compreender alguns dos problemas decorrentes das tecnologias contemporâneas.

Inicialmente a relação ser humano/máquina se apresentava com um cunho mais teórico e mitológico. Nascimento (2006) cita como exemplo o mito de Pigmalião, a de Ammon em Tebas, os oráculos, entre outros. A autora se refere às concepções aristotélicas de movimentos animais e mecânicos para discutir a analogia entre seres vivos e autômatos. Nascimento ressalta que:

> Como máquinas automáticas, têm a capacidade de autorregulação segundo uma finalidade predeterminada e em diferentes graus conforme a sua complexidade, mas, enquanto autômatos expressam, sobretudo uma mimetização dos movimentos do ser vivo, homem ou animal, que lhes confere uma aparência vitalista de autonomia orgânica. (Nascimento, 2006, p. 1034).

Nascimento (2006) enfatiza, ainda, que a criação de autômatos tinha como "modelo" as características dos animais em geral. Em contraposição, para problematizar o conceito de autonomia, a autora se apropria das características orgânicas e inorgânicas dos sistemas artificiais e naturais respectivamente.

Foram nos períodos renascentista e moderno que, segundo Nascimento (2006), foram obtidos resultados técnicos significativos. Esses períodos foram marcados, inicialmente, pelos pássaros autômatos bizantinos, os *jacquemarts* (que batiam as horas nas torres das igrejas) e, posteriormente, pelos vários mecanismos autômatos criados por Jacques de Vaucanson (1709-1782) como o seu tocador de flauta, de tamboril, de gaita e o célebre pato mecânico.

No período moderno, Nascimento (2006) enfatiza a influência do mecanicismo cartesiano segundo o qual a diferença entre o corpo dos animais e as máquinas artificiais é apenas relativa ao material e tamanho de suas partes, mas não em princípio. Ressalta Nascimento que para Descartes animais em geral, inclusive os humanos, e os autômatos possuem "[...] corpos com mecanismos predefinidos que contêm o seu próprio princípio de movimento, de vida e de morte" (Nascimento, 2006, p.1036). Todavia, ao diferenciar o ser humano dos demais animais-máquinas por ser dotada de uma razão (ou alma), substância pensante, imaterial e não sujeita às leis físicas, Nascimento aponta que "[...] Descartes acaba por concluir numa visão dualista de diferenciação entre ser humano e autômato, entre ser humano e animal, ao colocar a razão humana acima de qualquer recriação possível" (Nascimento, 2006, p. 1036).

É no contexto industrial emergente que Nascimento (2006) afirma ter sido inventada a primeira máquina automática programável, o tear de Joseph-Marie Jacquard (1801). Este abriu caminhos para que outras máquinas fossem desenvolvidas, de modo substituía substituir e auxiliar (de maneira eficiente) o processo produtivo realizado até então exclusivamente por seres humanos. Segundo Siegwart e Nourbakhs (2004), neste mesmo período os primeiros braços mecânicos fizeram grande sucesso, entretanto sofriam a desvantagem de serem fixos. Este autor trata em sua obra a mobilidade dos robôs sem a necessidade de supervisão. Nessa perspectiva, os autores apontam que um dos maiores desafios contemporâneos é lidar com a diversidade e as situações não controladas dos mais vários contextos ambientais.

Assim, na contemporaneidade as máquinas foram sendo aprimoradas e se desenvolveram a ponto de ser problematizado o conceito de autonomia para os hoje denominados sistemas artificias. Sistemas automáticos, como o termostato, não causam muitos debatesno âmbito da discussão sobre a agência, embora sejam sistemas que funcionem sem controladores humanos; mas outros sistemas artificiais hoje existentes, como o *Taranis* (2010), avião com finalidades bélicas e de reconhecimento que não é controlado remotamente por seres humanos, suscitam muitas questões sobre se possuem, ou não, algum tipo de agência artificial. Levando em consideração as possibilidades de ação de sistemas como o Taranis, *a* Royal Academy of Engineering (2009) analisa e discute os **graus de controle** dos seres humanos sobre os sistemas artificiais em relação à necessidade da intervenção humana para seu funcionamento. Assim, a Royal Academy of Engineering (2009) enumera quatro níveis:

> • sistemas controlados: em que os seres humanos têm controle total ou parcial, assim como os carros comuns;
>
> • sistemas supervisionados: fazem o que foi instruído pelo operador, assim como um torno mecânico ou outras máquinas industriais;
>
> • sistemas automáticos: que realizam funções fixas sem a intervenção de um operador, assim como um elevador;
>
> • sistemas autônomos: que são adaptativos aprendem e podem tomar "decisões". (The Royal Academy of Engineering, 2009, p. 02).[21]

[21]Controlled systems: where humans have full or partial control, such as an ordinary car; supervised systems: which do what an operator has instructed, such as a programmed lathe or other industrial machinery; automatic systems: that carry out fixed functions without the intervention of an operator, such as an elevator; autonomous systems that are adaptive, learn and can make ‘decisions’.

Os debates gerados em torno dos sistemas artificiais autônomos se dão devido à noção de que tais agentes tomam "decisões", porém não se sabe ao certo quem responderá pelas consequências de tais decisões se o programador, o fabricante, o usuário ou outro ser humano que tem algum vínculo com o sistema. A *Royal Academy of Engineering* (2009) se utiliza do argumento de que se deve adequar o grau de autonomia às tarefas desenvolvidas pelos sistemas artificiais. Além disso, defende que deve prevalecer o bom senso na atribuição de tarefas a tais sistemas, havendo aquelas que talvez devam ser confiadas somente a um ser humano. Em contrapartida há situações especialmente perigosas ou estressantes em que o ser humano pode não ter as informações contextuais para tomar a melhor decisão ou fazê-lo com a agilidade requerida, como, por exemplo, em um grande incêndio com baixa visibilidade.

Além do aspecto ético, existe também um aspecto legal. Assim como um ser humano, um sistema artificial também é passível de falha, embora o seja em diferente grau e possivelmente de modo diverso das falhas humanas. E se, por acaso um ser humano for gravemente ferido ou morto por falha de um sistema artificial? Alguém terá que ser responsabilizado. Entretanto, segundo a *Royal Academy of Engineering* (2009) nem todas as falhas humanas são responsabilizadas e que se provado que um sistema artificial pode ter uma porcentagem menor de erro do que um ser humano, em um procedimento cirúrgico, por exemplo, então seria até "prudente" delegar tal dever ao sistema artificial autônomo.Mais ainda, a *Royal Academy of Engineering* (2009) aponta um problema significativo: o caso, por exemplo, de sistemas autônomos que já apresentam um maior grau de eficiência e confiabilidade que os seres humanos na realização de tarefas corriqueiras, como é o caso da condução de um veículo em vias públicas, situação em que sistemas autônomos parecem estar sendo bem sucedidos. Considerando que, segundo alerta o Observatório Nacional de Segurança Viária[22], 90% dos acidentes de veículos no Brasil, por exemplo, são provocados por falhas humanas (como desatenção dos condutores, imprudência, negligência e desrespeito à legislação de trânsito), se veículos autônomos apresentarem um índice menor de falhas em experimentos exaustivos em situações de trânsito nas mais variadas condições, torna-se muito difícil argumentar contrariamente a sua utilização corriqueira e generalizada.

[22]Conforme informações disponíveis em 23/05/2016 no endereço eletrônico: http://www.onsv.org.br/noticias/90-dos-acidentes-sao-causados-por-falhas-humanas-alerta-observatorio/

Outra discussão relacionada à autonomia em sistemas artificiais gira em torno da capacidade de aprendizado de tais sistemas. Augusto (2007) relaciona a autonomia à capacidade do sistema interagir com o ambiente de maneira dinâmica, adaptando-se aos obstáculos e não realizando tarefas meramente repetitivas. O autor ainda frisa a importância da aprendizagem para que o sistema consiga desenvolver melhor esse comportamento autônomo. Entretanto, a aprendizagem na robótica se desenvolve sobre a problemática: "Quais são os tipos de conhecimentos desejáveis de serem adquiridos por um sistema artificial?".

Para observar a relação e consequência da aprendizagem e da autonomia em um sistema artificial, segundo Augusto (2007), é necessária à elaboração de um sistema real, embora as simulações computacionais auxiliem na validação de conceitos e na obtenção de resultados preliminares. A discussão sobre a aprendizagem para agentes artificias é assunto suficiente para outro trabalho, nos ateremos aqui a discussão da autonomia.

De um modo bem geral, Augusto (2007) traz o conceito de autonomia estritamente relacionado à concepção de "capacidade de ação independente". Ainda que seja necessário analisar a capacidade de reagir a estímulos externos e tomar decisões a curto prazo para que os sistemas artificiais possam ser considerados autônomos. Augusto (2007) retrata que em uma concepção de autonomia mais exigente é necessário que o sistema artificial possa tomar decisões a longo prazo, demostrando a capacidade de adaptar-se e mudar a sua relação com o meio. Desse modo, pode-se, então, definir autonomia de um sistema artificial, de forma ampla, como a habilidade de sensoriar seu ambiente e atuar correspondentemente em uma dada situação de forma apropriada, sem uma intervenção humana, em busca de seu objetivo ou na execução de uma tarefa.

Deve-se observar, entretanto, que, não existindo autonomia absoluta (nem mesmo para agentes humanos!), o que podemos fazer razoavelmente é fazer comparações entre sistemas e tentar estabelecer se um sistema é "mais autônomo" do que outro (Augusto, 2007, p. 9). A partir desta colocação, o autor levanta a questão de quais seriam então as capacidades mínimas para considerar um sistema autônomo e enumera alguns itens imprescindíveis do agente artificial em relação ao nicho de aplicações:

- Mover- se ao ambiente sem colidir com obstáculos;

- detectar sua própria localização no espaço;
- reagir a ambientes dinâmicos;
- detectar certos tipos de objetos;
- resolver questões decorrentes de haver objetivos conflitantes.

O debate sobre autonomia é atualmente um debate interdisciplinar e Haselager (2005) se propõe a diferenciar tal conceito analisado sob as perspectivas da robótica e da filosofia. Aprofundaremos melhor as ideias de Haselager (2005) na próxima seção.

3.2- Contribuições de Willem Haselager para o debate sobre a autonomia em sistemas artificiais

Visando aprofundar nossa discussão sobre autonomia em sistemas artificiais, encontramos no texto de Haselager (2005) subsídios e esclarecimentos de tal conceito. Primeiramente, o autor atribui igual importância às áreas da filosofia e da robótica para a análise do conceito de autonomia. A filosofia contribui apontando quais são as condições específicas e quais os parâmetros necessários para atribuir autonomia a um sistema. E a robótica contribui trazendo exemplos concretos e desafiadores para que a filosofia consiga ter mais subsídios para classificar o que é significativo ou não para atribuir autonomia a um sistema.

Para iniciar o debate sobre autonomia em sistemas artificiais, Haselager (2005) se fundamenta no contexto histórico da Inteligência artificial (IA) em que os sistemas autônomos artificiais passaram de *softwares* para robôs sofisticados que, por exemplo, viajam para Marte e são capazes de, sem a interferência de controladores humanos remotos, realizar tarefas complexas. O autor ainda acrescenta mais duas razões pelas quais os sistemas artificiais parecem ser possuidores de autonomia:

a. porque os robôs tem um corpo físico que permite ações efetivas, no mundo real e não apenas o virtual;
b. Porque o comportamento do robô conta com a característica “surpresa”, originalidade, pois seu comportamento é baseado em sua história de relação com o ambiente, ou seja, pode aprender em vez de realizar somente ações previamente determinadas em sua programação.

Considerando estes dois aspectos de sistemas artificiais robóticos, poder-se-ia dizer que o robô estaria *agindo* em um sentido bastante semelhante ao sentido em que se diz que seres humanos ou outros agentes naturais agem. No entanto, a possibilidade levantada quanto à agência de sistemas robóticos é considerada problemática, pois pode haver quem defenda que o corpo do robô, o sistema de controle, as adaptações e aprendizado são resultados de um comportamento muito dependente da programação e do *designer* humano para que se possa falar em autonomia e agência de sistemas artificiais.

Além desse debate é interessante frisar, com Haselager (2005), que os conceitos de autonomia e agência são muito obscuros, assim como outros conceitos centrais para o debate sobre as capacidades de sistemas artificiais .Por isso, Haselager (2005) propõe que as definições consideradas mantenham algum denominador comum e evitem concepções extremadas: tanto a de atribuir autonomia e agência a todos os sistemas, como o termostato, ou então de restringir demasiadamente sua atribuição, entendendo que apenas seres humanos adultos são agentes autônomos. Nesse viés, Haselager (2005) atenta para a diversidade da significação de autonomia e é justamente essa relação de experiência empírica da robótica, com a investigação conceitual da filosofia que enriquece e fortalece o debate sobre a legitimidade, ou não, da atribuição de autonomia e agência a sistemas artificiais.

Para fins de entendimento Haselager (2005) traz uma síntese de suas leituras sobre as definições de autonomia na perspectiva dos estudiosos da Inteligência artificial (IA). Desta maneira o autor conclui que "agentes autônomos operam em todas as condições razoáveis, sem recorrerem a um *designer*, operador ou controlador externo enquanto lidam com eventos imprevisíveis em um ambiente ou nicho[23]" (Haselager, 2005, p. 518). E complementa que não se deve recorrer a ajuda externa durante a ação, ou seja, o agente autônomo não é nem completamente independente, nem completamente dependente. Ao recorrer ao termo "condições razoáveis" indica que os sistemas artificiais autônomos tampouco poderiam ser previamente limitados em suas funções. Desta maneira se mediria o grau de autonomia de um sistema de acordo com o

[23] "Autonomous agents operate under all reasonable conditions without recourse to an outside designer, operator or controller while handling unpredictable events in an environment or niche".

tempo em que o sistema artificial interage com um ambiente não controlado realizando funções complexas com sucesso sem intervenções e supervisões humanas.

Em contraposição, Haselager (2005) nos lembra de que a perspectiva filosófica da autonomia enfatiza as razões pelas quais o agente está agindo, o objetivo que escolheu e como tal objetivo será alcançado. O termo autonomia está, para os filósofos, entrelaçado à noção de autogovernar-se, o poder de dar a si mesmo a própria lei, *autós* (por si mesmo) e *nomos* (lei/regra). Logo, o conceito de autonomia na perspectiva filosófica está profundamente relacionado à capacidade de um organismo agir por si, fazer suas escolhas e buscar seus objetivos, em vez de seguir os objetivos delimitados por outros agentes. Embora a história da filosofia seja complexa e apresente diversas noções de autonomia e agência, Haselager (2005) optou por apresentar uma visão mais platônica e aristotélica.

Baseado no contexto histórico do desenvolvimento da robótica Haselager (2005) constata que os sistemas automáticos foram desenvolvidos buscando alcançar uma menor dependência humana, ou seja, o ser humano não precisaria ficar constantemente operando-os, pois tais sistemas conseguiam repetir uma sequência de ações constante que poderia ser interrompida pelo ser humano a qualquer momento. Neste caso, segundo o autor, sistemas automáticos podem ser considerados um semiautônomos. O grau da autonomia do sistema aumentaria na medida inversamente proporcional ao grau de intervenção humana que o sistema necessita para operar com funcionalidade. Assim, o robô deve ir além das manipulações remotas, ampliar suas aptidões sensório-motoras e desenvolver a capacidade de agir em diversas circunstâncias e eventos sem que seja necessária alguma intervenção humana.

Mais ainda, Haselager (2005) afirma que é possível atribuir agência ao robô que atue autonomamente, pois é inadequado reduzi-lo à condição de evento ou objeto de que tratamos no primeiro capítulo, como no caso das gotas de chuva ou de uma pedra que rola por uma montanha. Os robôs portadores de agência, além de reagirem diante das condições ambientais, também são proativos na procura dos objetivos que estão ativos neles e tem um grau de possibilidade de escolha no sentido de poderem selecionar estratégias para alcançar seus objetivos.

Todavia há debates na perspectiva da filosofia sobre a natureza e grau de tal autonomia, pois, embora seja possível afirmar que o robô consiga agir de maneira

independente e ainda que tenha possibilidade de escolha de como agir para alcançar seu objetivo, o objetivo é previamente programado por seres humanos. Além disso, pode ser problematizada também a veracidade de que aumenta o grau de autonomia conforme diminui o envolvimento humano nas operações do robô, porque há uma quantidade considerável de trabalho humano *off-line* como o dos programadores e *designers*.

Mesmo levando em conta tais críticas, Haselager (2005) tem um posicionamento bem visível sobre a natureza da autonomia. Para o autor, o corpo é um fator muito importante a considerar quando se discute a atribuição de autonomia e agência. Haselager sugere que:

> A autonomia se baseia na formação de padrões de ação que resultam na automanutenção do sistema incorporado e se desenvolve durante a interação incorporada de um sistema com o seu ambiente. Há dois aspectos envolvidos nesta sugestão que necessitam mais pesquisa no âmbito da robótica. Em primeiro lugar, há a questão da integração entre o sistema de controle e o corpo. Em segundo lugar, a noção de homeostase merece um olhar mais atento. (Haselager, 2005, p.523).[24]

Haselager se reporta aos estudos sobre a agência de animais não humanos realizados por (Jakob Von Uexküll apud: Haselager, 2005) para quem o corpo biológico se caracteriza pela interação e a operação de todos os componentes que constituem o organismo. Von Uexküll, na condição de biólogo, entende, inclusive, que, devido à complexidade e capacidades de todo tipo que os caracterizam, os animais não humanos não podem mais ser considerados como meras máquinas (como ocorre na perspectiva do cartesianismo), mas como sujeitos que têm a percepção e a ação como atividades essenciais. Desta forma, Haselager (2005) acredita que o maior desafio da robótica seja a "produção do operador dentro do corpo" [25] dos robôs (Haselager, 2005, p. 524), principalmente no que tange ao estabelecimento da relevância de cada uma de suas metas nos diversos contextos. Tal capacidade é algo, lembra Haselager (2005), que os organismos desenvolveram ao longo de processos evolucionários por seleção natural.

---

[24] Autonomy is grounded in the formation of action patterns that result in the self-maintenance of the embodied system and it develops during the embodied interaction of a system with its environment. There are two aspects involved in this suggestion that bear some further investigation in relation to robotics. First of all, there is the issue of the integration between the control system and the body. Secondly, the notion of homeostasis deserves a closer look.

[25] "Building the operator into the body".

Entretanto, uma abordagem da robótica busca incorporar processos evolucionários em seus modelos. Trata-se da robótica evolutiva (*Evolution army robotics*) que, segundo (Nolfi apud: Haselager, 2005) trata de desenvolver nos robôs sistemas sensório-motores em um designer automático que envolveria a evolução artificial. "Evolução artificial envolve a utilização de algoritmos genéticos. Os 'genótipos' de robôs são representados pelos *bits* que podem codificar suas características morfológicas, bem como as características (tais como pesos e conexões de rede neural) dos seus sistemas de controle" (Haselager, 2005, p. 524) [26].

O tema da evolução artificial se torna interessante para o debate em torno da agência artificial, uma vez que Haselager (2005) afirma que a organização dos sistemas evolutivos é resultado de um processo auto-organizado, ou seja, seu comportamento emerge de suas interações com o ambiente, ainda que caiba aos seres humanos, tais como aos programadores e *designers*, selecionar os genes a serem cruzados para o desenvolvimento de um novo sistema artificial. Para isso muitos fatores devem ser levados em conta a iniciar pelo mapeamento genótipo e fenótipo.

Em suma, para Haselager (2015), o problema da autonomia em sistemas artificiais estaria fundamentado no como seriam determinados e em que se baseariam os objetivos de tais sistemas. O autor sugere que este embasamento estaria fundado na interação de todos os componentes corporais com o propósito orientado na homeostase. Assim, confirmaria a relevância das abordagens tanto da filosofia quanto da robótica para a discussão de autonomia de sistemas artificiais.

É importante enfatizar que a homeostase está relacionada à noção de automanutenção que até então é uma característica presente apenas nos organismos vivos, ou seja, o corpo tende a se manter em um estado global compatível com sua funcionalidade, é o que acontece, por exemplo, quando há uma concentração demasiada de glicose no sangue. Os receptores do pâncreas captam essa informação e iniciam um processo para liberar a insulina e assim diminuir a concentração de glicose no sangue.

Após concluir a reflexão sobre autonomia para sistemas artificiais na perspectiva de Haselager (2005), nos propomos a apresentar uma reflexão especulativa deste tema na perspectiva de Hans Jonas, pois encontramos entre ambas as perspectivas

[26] Artificial evolution involves the use of genetic algorithms. The 'genotypes' of robots are represented as bits that can cod their morphological features as well as the characteristics (such as weights and connections of neural network) of their control systems.

alguns pontos confluentes. O mais relevante deles diz respeito à centralidade do tema da vida, sua manutenção e preservação, na determinação de um campo ético fundado na responsabilidade de seres autônomos. Como ele ressalta: "[...] forma – isto é, forma autônoma, em si real – é um caráter essencial da vida". (Jonas, 2004, p.102).

3.3- Especulações acerca da atribuição de autonomia e agência a sistemas artificiais em uma perspectiva jonasiana

Nesta seção iremos investigar alguns aspectos da filosofia jonasiana no que tange à vida, aos organismos e a agência que tratam de questões semelhantes às abordadas por Haselager (2005) em seu estudo da noção de autonomia. Em especial, a tese de que a autonomia está entrelaçada à corporeidade dos sistemas, sobretudo à auto manutenção e à busca e priorização de objetivos, e não apenas a aspectos racionais ou reflexivos. Entendemos que, preservadas as diferenças entre ambas as concepções, a perspectiva ética desenvolvida por Jonas (2004) também ressalta que a autonomia também está relacionada ao organismo vivo, comportando aspectos corpóreos e ecológicos, além dos valorativos. Cabe ressaltar que no pensamento jonasiano, assim como em parte significativa da tradição da filosofia, os termos autonomia e liberdade aparecem correlacionados, mais, segundo Jonas (2004) a autonomia é um elemento da do agir livre ou, em nosso vocabulário, da agência. A liberdade não está presente somente no ser humano, segundo Jonas (2004), como também em todo o mundo orgânico. Ela consiste

> [...] em certa interdependência da forma com relação a sua própria matéria [...] e o aumento desta independência ou liberdade é o princípio de todo o progresso na história da evolução da vida, que em seu decurso apresenta outras revoluções, cada uma delas um novo passo na direção tomada, isto é, cada um abre um novo horizonte de liberdade. (Hans Jonas, 2004, p. 104)

Consideramos que esta tese jonasianas e aproxima da noção de autonomia enquanto "busca de objetivos" que Haselager (2005) entende ser um dos componentes necessários para que um sistema artificial possa ser considerado autônomo e, assim sendo, portador de agência. Outro ponto de confluência entra ambos diz respeito à recusa de uma concepção puramente intelectualista da autonomia: para Haselager (2005) processos cognitivos em geral, inclusive os deliberativos relacionados a tomadas

de decisão, envolvem componentes corporais e contextuais. Entendemos que Jonas endossaria esta tese com base no seguinte texto:

> Assim, a autonomia em relação à natureza, estabelecida e afirmada na auto causalidade do organismo – autonomia que não é mecânica – tem seu exato preço na dependência existencial em relação à natureza, que é totalmente estranha à estabilidade do ser da matéria sem vida. (Hans Jonas, 2004, p. 123)

Cabe destacar que na passagem acima, Jonas ressalta que a autonomia afirmada na auto causalidade, própria dos seres vivos, não é mecânica. No entanto, sistemas artificiais capazes de aprender e de auto organizar-se, como ressaltado por Haselager (Haselager, 2005, p. 524), não são mecânicos no sentido de estarem inteiramente determinados por uma programação prévia. Em outras palavras, tais sistemas artificiais estão sujeitos às leis naturais e às funcionalidades previamente programadas, mas têm a possibilidade de tomarem decisões. Podemos nos perguntar: esta caracterização não descreve adequadamente também os sistemas naturais ou seres vivos, especialmente quando consideramos seus aspectos genéticos? Será que há mesmo uma diferença ontológica entre sistemas artificiais e naturais?

Além desse aspecto, discussões levantadas por Jonas (2004) sobre a natureza da vida e sua relação com a matéria de que os seres vivos são compostos podem nos auxiliar a compreender mais uma possível aproximação entre as teses jonasianas e a abordagem cognitivista defendida por Haselager (2005).

As reflexões de Jonas consideram de que modo a vida e sua relação com a morte foi inicialmente concebida por várias culturas humanas, para as quais a vida seria uma propriedade geral da natureza associada ao movimento e não apenas dos organismos (constituindo um panvitalismo). De acordo com esta concepção, a vida seria uma espécie de princípio ou "regra geral" enquanto a morte constituiria uma exceção.

Nessa perspectiva, a natureza da morte constitui uma questão central. Ressalta Jonas: "Na medida em que a vida era considerada como um estado primário das coisas, a morte destaca-se como o enigma que perturba". (2004, p.18). Desse modo, vida e morte são opostas. Vida é natural, que se pode compreender, ela é a regra. A morte, entretanto, é sua oposição, é o não natural, o não compreendido. É a partir dessas

inquietações sobre a natureza da morte que possivelmente tenham surgido inquietações metafísicas.

Já no pensamento moderno ocidental, iniciado no Renascimento, ocorre o inverso, a morte é o natural e o problema é a vida. A natureza passa a ser concebida como algo desprovido de vida, sendo que a matéria que a constitui é considerada muito simples para instanciá-la por si mesma. Jonas salienta que: "Em consequência, o que agora exige uma explicação no universo [...]é a existência da vida, e esta explicação tem que ser dada em termos da matéria inerte." (Jonas, 2004, p. 20). Neste período o questionamento que se faz é: Como existir vida em um universo formado de pura matéria? Jonas aponta que

> Só na morte é que o corpo deixa de ser um enigma: na morte ele retorna do comportamento enigmático e inortodoxo da vida para o estado claro e 'familiar' de um corpo dentro do conjunto corporal, cujas leis gerais constituem a regra de toda compreensão (Jonas, 2004 p. 21-22).

Neste caso, Jonas (2004) enfatiza que o termo morte se aplica ao corpo, constituído de componentes materiais, e que só pode ser aplicado ao que pode ser ou já foi vivo. Em contraposição "[...] vida quer dizer vida material, portanto corpo vivo, em suma, ser orgânico. No corpo está amarrado o nó do ser, que o dualismo rompe, mas não desata" (Jonas, 2004, p. 34). Além disso, a vida, segundo Oliveira "[...] é compreendida [por Jonas] a partir de um 'si mesmo' que ao mesmo tempo remete à identidade interior de cada vivente e ao seu isolamento dos demais [...] (Oliveira, 2011 p. 47)".

Na contemporaneidade, nos atrevemos a dizer que com o desenvolvimento das novas tecnologias, principalmente as relacionadas a saúde, reconhecemos como problemáticos os dois conceitos, o de vida e o de morte, especialmente quando nos deparamos com situações limite: um ser humano em estado vegetativo podemos legitimamente dizer que está vivo? Um sistema artificial capaz de aprender podemos legitimamente dizer que está morto?

Outro fator relacionado ao corpo e discutido por Jonas (2004) tratada natureza da matéria, "dividida" em orgânica e inorgânica. No que tange a matéria inorgânica, o autor afirma que ela pode ser identificada a partir de sua posição no espaço e tempo. Ela

também traz como característica ser igual a ela mesma (auto identidade), pois ela é sempre constante e contínua nas suas dimensões de espaço-tempo.

No que tange à identidade orgânica, Jonas (2004) enfatiza a necessidade de considerar a "continuidade metabólica da forma orgânica" (Jonas 2004, p.105) que é um dos fatores diferenciadores da matéria não orgânica. Por meio dos processos metabólicos, o não-vivo (como sais minerais ingeridas em sucos de frutas, por exemplo), passa a integrar o vivo (quando esses sais minerais passam a compor os tecidos do organismo). Desse modo há uma transformação daquela matéria com identidade fixa (sem vida) para uma matéria com uma identidade que pode ser transmitida e replicada(via processos reprodutivos). Ela é uma *identidade interior* e que só pode ser observada e ter suas consequências deduzidas.

> O observador da vida tem que estar preparado através da vida. Noutras palavras, dele se exige o ser orgânico com sua experiência própria, para que esteja em condições de deduzir aquela 'consequência' que de fato ele tira continuamente e esta é a vantagem, tão teimosamente negada ou caluniada, da história da teoria do conhecimento – **a vantagem de termos um corpo, ou de sermos um corpo**. Em suma, nós estamos preparados por aquilo que somos. Só por meio da interpolação da identidade interior, que assim se torna possível, é que o fato meramente morfológico (e como tal carente de sentido) da continuidade metabólica é compreendido como ato incessante, isto é, a continuidade é compreendida como autocontinuação. (Jonas, 2004, p. 105, destaque nosso).

Desta maneira, a agência e autonomia atribuída somente aos organismos vivos envolvem aspectos dinâmicos que dizem respeito não apenas a capacidades dos agentes individualmente (como a dinâmica metabólica), mas a capacidades que envolvem a perpetuação de um tipo de organização (ou espécie) ao longo do tempo. Novamente, nos deparamos com uma situação problemática que impede o estabelecimento de limites claros entre sistemas naturais e artificiais no que diz respeito a poderem ser considerados como evidentemente autônomos ou carentes de autonomia. Isto porque não há um impedimento de princípio que impeça haver sistemas artificiais capazes de se autorreplicarem e perpetuem sua linhagem (programação e conhecimento adquirido) de modo análogo ao dos seres vivos.

Assim, considerando que Jonas (2004) ainda afirma que a identidade orgânica é "uma identidade que se faz de momento a momento, que sempre de novo se afirma forçando as forças igualizadoras da mesmidade física ...". (2004, p.105-106), e

considerando as capacidades inovadoras das novas tecnologias, que incluem a aprendizagem, a tomada de decisão e a capacidade de automanutenção e podem vir a incluir a autorreplicação, parece difícil encontrar um argumento consistente contrário à atribuição de, pelo menos, uma proto agência a sistemas artificiais. Se assim for, precisamos considerar a possibilidade de atribuir também algum grau de responsabilidade a tais agentes, pelo menos de corresponsabilidade pelas possíveis implicações de seu agir no mundo, o que exige uma revisão radical de concepções éticas e jurídicas profundamente enraizadas na cultura ocidental em torno às noções de agência moral e responsabilidade..

Considerações finais

Tendo esta dissertação o objetivo de analisar os conceitos de autonomia e responsabilidade no contexto tecnológico contemporâneo capaz de produzir sistemas artificiais dotados de um alto grau de complexidade, inicialmente levamos em consideração o que se entende por "ação" e, consequentemente, por "agência" no primeiro capítulo desta dissertação.

Inicialmente foram abordadas as distinções entre as noções de evento e objeto e entre objetos causais e agentes causais, inspiradas especialmente nos trabalhos de Alícia Juarrero (1999), o que nos levou a um embate: se a distinção entre objeto e evento é obscura, visto que ambos parecem relacionados e até interdependentes, como poderíamos, então, distinguir agentes causais de objetos causais, se a mesma ambiguidade se coloca em relação a estes últimos?

Para discutir essa questão buscamos investigar a diferenciação proposta por teóricos da ação entre uma ação básica (que pressupõe intenção) e uma não básica, tarefa essa igualmente difícil de alcançar. Ainda, apresentamos a crítica proposta por Frankfurt (1978) às teorias causais da ação, pois para ele as ações não podem ser explicadas a partir de causas antecedentes, embora não negue que as ações tenham causas, de modo que é mais válido recorrer as ocorrências que estão presentes na ação analisada. Se por um lado a ação intencional, em uma visão mais tradicional, é vista como característica do ser humano, por outro lado Schlosser (2015) apresenta o fazer ativo e a agência, subdivididos em níveis, não havendo necessidade da racionalidade estritamente humana em todos os níveis da agência, o que possibilita incluir organismos não humanos e sistemas artificiais como possíveis portadores de agência.

Considerando que as noções de autonomia e agência estão intimamente relacionadas com a de responsabilidade, uma vez que se considera que um ser autônomo é tal por possuir a capacidade de tomar decisões e assumir as implicações que decorrem elas, no capítulo 2 investigamos o conceito de responsabilidade. Para isso, optamos por recorrer à noção de responsabilidade proposta por Hans Jonas (2004, 2013), por ele ter tido especial interesse em discutir o conceito de responsabilidade no contexto da tecnologia.

Assim, no segundo capítulo analisamos e discutimos teses jonasianas sobre a relação entre tecnologia e responsabilidade. Primeiramente, analisamos a concepção

jonasiana de desenvolvimento tecnológico e como ele se apresenta de modo diferente ao longo da história humana, passando por cinco estágios principais (mecânico, químico, elétrico, eletrônico e biológico) aos quais somamos um sexto, o da Inteligência Artificial no sentido forte, capaz de produzir sistemas artificiais autônomos em algum grau não trivial. Destacamos, ainda, que para Hans Jonas, é justamente a capacidade de desenvolvimento da tecnologia, a qual está afetando significativamente a dinâmica ecológica do planeta, que impõe ao ser humano uma responsabilidade não apenas pelas ações presentes, mas pelas possíveis implicações futuras do uso generalizado dessas tecnologias, especialmente as contemporâneas. Para dar conta dessa responsabilidade ampliada para as possíveis consequências futuras das ações atuais, Jonas (2004) propõe um método reflexivo, denominado "heurística do temor", uma generalização e radicalização do princípio da prudência. O temor diante de possíveis consequências negativas das tecnologias para as gerações futuras, mesmo que remotas, deve ser o "guia" para as escolhas humanas no que tange, sobretudo, à sua criação e utilização.

Por fim, o terceiro capítulo trata mais especificamente da noção de autonomia e agência em sistemas artificiais. Em um primeiro momento, procuramos caracterizar o conceito clássico de autômato e alguns de seus exemplos históricos. Em seguida, analisamos a distinção entre sistemas artificiais proposta pela *Royal Academy of Engineering* entre sistemas automáticos e sistemas dotados de algum grau de autonomia na medida em que têm a capacidade de aprender e, a partir da aprendizagem, tomar decisões não constantes em sua programação inicial. Em seguida, apresentamos teses centrais da abordagem da cognição incorporada e situada adotada por Willem Haselager (2005) para a discussão da noção de sistemas artificiais autônomos. A partir da perspectiva incorporada e situada da cognição, Haselager discute especialmente a legitimidade de atribuição de autonomia a sistemas robóticos situados e incorporados de modo análogo aos organismos naturais, considerando suas potencialidades de ação no mundo. Para Haselager, o limite entre o vivo e o não vivo, a vida natural e a vida artificial, mais claramente determinado quando se tratava apenas de seres automáticos, torna-se atualmente cada vez mais difuso e nebuloso devido às novas possibilidades de aprendizagem e auto-manutenção de que são dotados alguns sistemas artificiais. Porém, nesse contexto, devemos lembrar que, na esteira das colocações de Hans Jonas, os

sistemas artificiais não têm corpos com capacidade de emular os processos metabólicos próprios dos sistemas naturais e que consideramos desempenhar um papel central em sua ação autônoma. Tal característica seria primordial para atribuir a qualquer sistema a posse de agência, autonomia e, consequentemente, de responsabilidade. Haselager (2005) considera que os agentes artificiais não teriam, por princípio, o mesmo grau de autonomia de organismos vivos. Na esteira de Haselager (2005),concordamos que sistemas artificiais não possuem o mesmo grau de autonomia que possuem sistemas naturais, mas nem por isso os sistemas artificiais podem ser considerados simplesmente equivalentes a objetos quaisquer, como pedras e gotas de chuva (exemplos de elementos passivos e determinados nos eventos naturais). Consideramos que não cabe tal equivalência na medida em que sistemas artificiais sejam capazes de aprender, de constituir uma memória do curso de suas ações e suas consequências, e de tomar decisões levando em consideração tal memória.

Concluímos então que o termo "agente artificial" pode parecer problemático, a primeira vista, por serem os robôs ainda muito dependentes da supervisão humana, seja em sua produção, seja em sua manutenção ou no monitoramento remoto de suas performances. Todavia, também parece problemático recusar atribuir a sistemas artificiais capazes de aprender, tomar decisões e cuidar de vários aspectos de sua manutenção um certo grau de agência, autonomia e até de responsabilidade pelas possíveis consequências (positivas e negativas) de suas performances no mundo.

Entendemos que cabe aos idealizadores e implementadores de sistemas artificiais levar em consideração a sugestão de Hans Jonas (2004) de ponderar as possíveis implicações a longo prazo da produção de agentes autônomos artificiais para as gerações futuras, especialmente para finalidades bélicas. Cabe aos filósofos tomar a iniciativa de estabelecer diálogos interdisciplinares com cientistas da computação para auxiliá-los nessas ponderações.

Consideramos que a heurística do temor proposta por Hans Jonas para avaliar as possíveis implicações a longo prazo das tecnologias contemporâneas constitui uma baliza moral para os seres humanos devido a sua identidade biológica e cultural, especialmente no que se refere às emoções relacionadas às gerações futuras. Nessa medida, cabe indagar se sistemas artificiais teriam condições de aplicar adequadamente

essa relevante ferramenta moral em benefício da preservação da natureza, incluindo os seres vivos.

Entretanto, consideramos importante ressaltar que, surpreendentemente, ao refletirmos em profundidade e despidos de preconceitos, especialmente intelectualistas, sobre os motivos pelos quais atribuímos a alguém agência e responsabilidade por suas ações, evitaremos equiparar certos sistemas artificiais a meras coisas, a objetos que pertencem à mesma categoria que pedras rolantes ou gotas de chuva.

Referências

ARENDT, Hannah. *Responsabilidade e julgamento.* Tradução de RosauraEinchenberg**.** São Paulo: Companhia das Letras, 2004.

AUGUSTO, Sergio Ribeiro. *Uma plataforma móvel para estudos de autonomia***.** 2007. 141 f. Tese (Doutorado) - Curso de Engenharia, Engenharia de Telecomunicações e Controle, Escola Politécnica da Universidade de São Paulo, São Paulo, 2007. Disponível em: <www.teses.usp.br/teses/disponiveis/3/.../teseSergio Revisada.pdf >. Acesso em: 02 nov. 2015.

BATTESTIN, Cláudia; GHIGGI, Gomercindo. O princípio responsabilidade de Hans Jonas:um princípio ético para os novos tempos. *Thaumazein,* Santa Maria, v. 06, p.69-85, out. 2010. Disponível em: <http://sites.unifra.br/Portals/1/ARTIGOS/numero_06/battestin_5.pdf>. Acessoem: 12 mar. 2014.

BOSTROM, Nick. *When Machines Outsmart Humans.* Futures**,** Oxford, v. 35, n. 7, p.759-764, nov. 2000. Disponível em: <http://www.nickbostrom.com/2050/outsmart.html>. Acesso em: 10 nov. 2015.

CARTWRIGHT, Jon. *Rise of the robots and the future of war*.The Guardian. Reino Unido, p. 128-129. 21 dez. 2010. Disponível em: <https://www.theguardian.com/technology/2010/nov/21/military-robots-autonomous-machines>. Acesso em: 15 maio 2015.

CARVALHO, Hélder Buenos Aires de. Uma filosofia para compreender a crise ambiental. *Instituto HumanitasUnisinos,* São Leopoldo - Rs, p.1-2, 29 ago. 2011. Entrevistado por: Márcia Junges. Disponível em: <http://www.ihuonline.unisinos.br/index.php?option=com_content&view=article&id=4036&secao=371>. Acesso em: 20 jun. 2015.

CASATI, Roberto; VARZI, Achille. *Events.* Califórnia: Edward N. Zalta, 2015. Disponível em: <https://plato.stanford.edu/archives/win2015/entries/events/>. Acesso em: 20 nov. 2015.

CHAROVA, Kseniya; SCHAEFFER, Cameron; GARRON, Lucas. *ArmyRobots.*2011. Disponível em: <https://cs.stanford.edu/people/eroberts/cs181/ projects/ 2010-11/ComputersMakingDecisions/army-robots/index.html>.Acesso em: 06 mar. 2016.

ESHLEMAN, Andrew. *Moral Responsibility*. Califórnia: Edward N. Zalta, 2016. Disponível em: <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=moral-responsibility>. Acessoem: 25 maio 2016.

FRANKFURT, Harry. *The problem of action.* Illinois: American Philosophical Quarterly, 1978.P.157-162 Disponívelem: <https://voices.uchicago.edu/wittgenstein/files/2007/10/frankfurt-the-problem-of-action.pdf>. Acessoem: 10 jan. 2016.

HASELAGER, Willem F.g. *Robotcs, philosophy and the problems of autonomy.* Usa: John Benjamins PublishingCompany, 2005. Disponível em: <http://www.socsci.ru.nl/haselag/publications/PragCogHaselager05.pdf>. Acesso em: 23 jun. 2016.

JONAS, Hans. *O princípio responsabilidade:* Ensaio de uma ética para a civilização tecnológica. Rio de Janeiro: Contraponto- PUCRJ, 2006. Tradução de:Marijane Lisboa eLuiz Barros Montez.

_____*O princípio vida:* fundamentos para uma biologia filosófica. Rio de Janeiro: Vozes, 2004. Tradução de: Carlos Almeida Pereira.P.132-158

_____Técnica, medicina e ética*:* sobre a prática do princípio responsabilidade. São Paulo: Paulus, 2013. Tradução de: Grupo de Trabalho Hans Jonas da ANPOF.
JUARRERO, Alicia. *Dynamics in Action:*Intentional Behavior as a Complex System. London: Massachusetts Institute of Technology, 1999.
LENMAN, James. *Reasons for Action:*Justification vs. Explanation**.** Califórnia: Stanford Encyclopedia Of Philosophy., 2009. Disponívelem: <https://stanford.library.sydney.edu.au/archives/spr2010/entries/reasons-just-vs-expl/>.Acesso em: 10 jan. 2015.
NASCIMENTO, Susana. Automatizações no inorgânico: aproximações ao estudo social de criaturas artificiais. *Análise Social,* Lisboa, v. 181, n. 221, p.1033-1056, set. 2006. Disponível em: <http://analisesocial.ics.ul.pt/documentos/1218723500Q7hZF9ni2Pf18AE4.pdf>. Acessoem: 21 jun. 2016.
NOORMAN, Merel, *Computing and Moral Responsibility.*Califórnia: Edward N. Zalta, 2014. Disponível em: <http://plato.stanford.edu/archives/sum2014/entries/computing-responsibility/>. Acesso em: 19 out. 2014.
OLIVEIRA, Jelson. . A heurística do temor e o despertar da responsabilidade. *Instituto HumanitasUnisinos*, São Leopoldo, v. 371, n. 11, p.1-3, 29 ago. 2011. Entrevistado por: Márcia Junges. Disponível em: <http://www.ihuonline.unisinos.br/index.php?option=com_content&view=article&id=4035&secao=371&limitstart=1>. Acesso em: 24 abr. 2015
_____ A transmalidade do homem: Uma premissa do Princípio Responsabilidade. In: SANTOS, Robson dos; OLIVEIRA, Jelson; ZANCANARO, Lourenço. *Ética para uma civilização tecnológica:* em diálogo com Hans Jonas. São Paulo: São Camilo, 2011. P. 41-60.
PIZZI, Jovino. Jonas e o enaltecimento da Heurística: A responsabilidade frente ao futuro ameaçado.SANTO, Robson dos; OLIVEIRA, Jelson; ZANCANARO. *Ética para uma civilização tecnológica:* em diálogo com Hans Jonas. São Paulo: São Camilo, 2011.P. 97-114.
PRADO, Marie del. *This drone is one of the most secretive weapons in the world.* 2015. Disponível em: <http://www.businessinsider.com/british-taranis-drone-first-autonomous-weapon-2015-9>. Acesso em: 25 fev. 2016.
REIS, Filipe de Abreu. *A Responsabilidade Civil.* 2012. Disponível em: <http://www.viajus.com.br/viajus.php?pagina=artigos&id=1227>. Acesso em: 27 jul. 2015.
SCHLOSSER, Markus, *Agency*. Califórnia: Edward N. Zalta, 2015. Disponível em <http://plato.stanford.edu/archives/fa ll2015/entries/agency/> Acesso em: 10 dez. 2015.
SEGERBERG, Krister; MEYER, John-Jules; KRACHT, Marcus.*The Logic of Action.*Califónia: Edward N. Zalta, 2009. Disponível em: <http://plato.stanford.edu/archives/win2013/entries/logic-action/> Acesso em:3 mar. 2016.
SGANZERLA, Anor. O sujeito ético em Hans Jonas : os fundamentos de uma ética para a civilização tecnológica., Robson dos; OLIVEIRA, Jelson; ZANCANARO, Lourenço. *Ética para uma civilização tecnológica***:** em diálogo com Hans Jonas. São Paulo: São Camilo, 2011.P.115-128.
SHIPLEY, Thomas F.. An invitation to a Event. In: SHIPLEY, Thomas F.; ZACKS, Jeffrey M.. *Understanding events:* From perception to action. New York: Oxford University Press, 2008. p. 3-60. Disponívelem: <https://books.google.com.br/books?id

=dOTATPNYQmgC&pg=PR9&hl=ptBR&source=gbs_selected_pages&cad=2#v=onepage&q&f;=false>. Acessoem: 03 out. 2015.
SIEGWART, Roland.NOURBAKHSH, Illah R. *Introduction to Autonomous Mobile Robots***.** London: A Bradford Book TheMit Press, 2004. p. 1-12. Disponívelem: <http://home.deib.polimi.it/gini/robot/docs/siegwart.pdf>. Acesso em: 02 ago. 2016.
SILVA, Luzia Gomes da. *Estudo da natureza jurídica e da responsabilidade civil por danos morais.* 2013. Disponível em: <http://www.conteudojuridico.com.br/?artigos&ver=2.42099&seo=1>. Acessoem: 18 ago. 2015.
SMILEY, Marion.*Collective Responsibility***.**Califórnia*:* Edward N. Zalta, 2011. Disponível em <http://plato.stanford.edu/archives/fall2011/entries/collective-responsibility/>. Acessoem: 09 mai. 2014.
THE ROYAL ACADEMY OF ENGINEERING.. *Autonomous Systems:* Social, Legal and Ethical Issues. London: The Royal Academy Of Engineering, 2009.
VIANA, Wellistony C. *Hans Jonas e a filosofia da Mente*. São Paulo: Paulus, 2006 P.119-131.