

AUTO-ILUSÃO E OS LIMITES INTERNOS DO NARRATIVISMO*

José Sérgio Duarte da Fonseca
UFRJ/FAPERJ

Resumo: O artigo tem como objetivo examinar duas versões antagônicas do conceito de pessoa, o narrativismo anti-reducionista, proposto por Alasdair MacIntyre, e o modelo reducionista, proposto por Derek Parfit. Será mostrado que nenhuma das duas explicações satisfazem dois requerimentos essenciais que devem ser observados por toda e qualquer teoria da pessoalidade, a saber, o de ser capaz de prover um critério fidedigno de distinção entre pessoas reais e ficções e o de poder explicar a irracionalidade (p.ex., a auto-ilusão). As consequências desta crítica são expostas na conclusão.

Palavras-chave: Identidade pessoal, irracionalidade, narrativismo, MacIntyre, Locke

Abstract: This paper examines two opposed views on the concept of person, the anti-reductionist narrativism proposed by Alasdair MacIntyre and the reductionist account proposed by Derek Parfit. I will argue that these two views do not satisfy two essential demands that must be observed by any theory of personal identity, namely: being able to providing a trustful criterion of the distinction between real persons and fictional characters and being able to explain irrationality (e.g., self-deception). The consequences of the criticism will be exposed in the conclusion.

Key words: Personal identity, irrationality, narrativism, MacIntyre, Locke

* Este artigo faz parte de uma pesquisa acolhida no Centro de Ética e Filosofia da Mente (Departamento de Filosofia da UFRJ), e realizada sob a vigência de uma Bolsa de Fixação de Pesquisador da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

1. MacIntyre, Parfit e a identidade pessoal

No capítulo XV de *Depois da virtude*, MacIntyre (MacIntyre [1985]) avalia e critica os obstáculos encontrados contemporaneamente por aqueles que, como ele, tentam resgatar a concepção pré-moderna, ou mais exatamente aristotélica de pessoa. Estes obstáculos dividiram-se em dois grupos, os sociais e os filosóficos. É importante notar que, para MacIntyre, a estrutura social deve ser levada sempre em conta para a avaliação da argumentação filosófica; assim sendo, esses dois obstáculos não devem ser considerados como mutuamente independentes. Os obstáculos sociais estão ligados à complexificação das estruturas sociais, característica da era moderna, e à decorrente ruptura da unidade da vida humana, tal como era supostamente experimentada em tempos pré-modernos, especialmente na Antigüidade. Contrapondo-se ao período anterior, a vida humana aparece agora completamente segmentada, seja em termos da distinção entre o público e privado, seja em termos do parcelamento da vida dos próprios indivíduos em partes estanques, como, p. ex., em infância e velhice ou em funções sociais compartimentadas e díspares, sem que haja conexões internas umas com as outras. Além disso, para complementar essa fragmentação, cada uma das partes da vida se definiria através de suas próprias normas e regras de comportamento. O segundo obstáculo ao projeto de reconstrução da ética aristotélica adviria das doutrinas filosóficas que estão em cena no debate contemporâneo, cujas teses teriam como efeito prático o obscurecimento da unidade da vida humana. Esse é o ponto onde MacIntyre se refere explicitamente, ainda que de passagem, a Parfit. Para que se torne mais clara a perspectiva através da qual MacIntyre critica Parfit, farei uma rápida digressão sobre os aspectos gerais de sua crítica à modernidade, já que os enganos imputados a Parfit adviriam exatamente da herança moderna que está no centro de sua concepção de *self*.

A visão teleológica da vida como busca pelo bem e pelo exercício das virtudes pressupõe a unidade desta vida, pois é ela que garantiria a racionalidade da vida moral. Como é sabido, a ética das virtudes possui um esquema triádico, i. e., o homem tal como é, o homem tal como poderia ser, se compreendesse sua natureza essencial (e seu *telos*) e os preceitos morais, que funcionam como um instrumento que permite a passagem de um estado de coisas ao outro, do “ser” ao “dever ser”. Este esquema conceitual foi posto em cheque na era moderna, que se caracteriza exatamente pela rejeição secular da teleologia católica e protestante, bem como pela rejeição filosófica e científica do aristotelismo. Cai por terra, assim, todo o conjunto que sustentava a concepção teleológica da vida humana, que se mantivera inalterada — “inalterada” nesse aspecto teleológico, mas não em outros pontos — desde a Antigüidade até a Idade Média, juntamente com sua estrutura triádica.

Essa transformação é percebida por MacIntyre como uma perda, isto é, como a deformação da linguagem da moralidade, que explicaria a desordem que caracterizaria, segundo ele, o debate moral contemporâneo. Para MacIntyre, o projeto Iluminista falha exatamente porque se baseia em um esquema moral aristotélico mutilado, “pós-diluviano”, resultante da eliminação de seu caráter teleológico. Essa é a catástrofe que, devido à sua deformidade constitutiva, a sensibilidade moderna não tem condições de perceber enquanto tal, ou seja, a sensibilidade moderna não tem condições de perceber o desmembramento do esquema aristotélico. Dada essa contextualização histórica, torna-se fácil entender que a filosofia, tanto a analítica quanto a fenomenológica, não poderia perceber tal desordem, pois elas mesmas pertencem a uma ordem social posterior à catástrofe, pela qual são em parte responsáveis.

Em vista disso, não seria mais possível fundamentar o acordo racional no debate moral contemporâneo, caracterizado, segundo MacIntyre, pela incomensurabilidade das diversas posições sobre a natureza da moralidade. Restaria apenas a aparência de racionalidade e de impessoalidade dos argumentos. O caos se deve à utilização de conceitos morais sem a devida contextualização de outrora. A fonte da confusão está no fato de que a discussão moral é tratada ao mesmo tempo como exercício racional e enunciação de juízos assertivos. Com isto, passa a circular a opinião de que o desacordo moral não pode ser resolvido, não devido a algum impedimento estrutural, explicável pelo nosso contexto histórico e social, mais sim devido a um aspecto necessário do próprio discurso moral. Segundo MacIntyre, essa opinião é muito semelhante à doutrina emotivista. O emotivismo como doutrina define-se por sua visão de que todo o julgamento moral não é mais do que expressão de preferência, refletindo atitudes e sentimentos, e cujas proposições não são, portanto, nem verdadeiras nem falsas. O emotivismo confundiria condições contingentes, isto é, a conjuntura histórica particular que a sustenta, o Iluminismo fruto da modernidade, com uma realidade universal, tornando-se a opinião corrente em nossa cultura. Conseqüentemente, uma vez diagnosticados todos os empecilhos de natureza histórico-cultural que impedem que o debate moral contemporâneo possa se dar de forma racional, o esforço de MacIntyre concentra-se na tentativa de colocar em ordem os fragmentos dispersos, com o objetivo de desenvolver uma proposta de narrativa histórica que reconstituísse a linguagem moral esquecida, devolvendo-lhe a devida racionalidade.

A partir desta rápida exposição, podemos compreender a perspectiva geral a partir da qual MacIntyre critica Parfit. O ponto central de sua crítica a Parfit é o de que a concepção reducionista de pessoa, em que o conceito de identidade pessoal é caracterizado em termos de uma relação específica entre seus elementos constituintes, isto é, crenças, desejos e ações, sem que seja necessário, dessa forma, fazer qualquer referência a “fatos adicionais”,

tornaria impossível que a este sejam atribuídas as virtudes aristotélicas, pois elas dependem da inserção prévia do *self* em uma comunidade de valores, valores esses que definem um padrão moral, que funciona como um critério unificador para a boa vida. Essa exigência de unidade de vida está ligada, como já foi dito, ao projeto macintyreano de resgate da ética das virtudes, que se caracteriza pela visão do homem como tendo uma essência que define sua finalidade. É a constituição teleológica de nossa vida que faz com que ela só tenha sentido se a entendermos como assumindo a forma de uma história a ser contada, sendo essa narrativa de vida a condição necessária para que conheçamos o que somos, bem como para saber no que nos tornamos e para onde vamos. Dessa forma, nossa compreensão do que seja a vida é necessariamente temporal, sendo que essa temporalidade não é meramente física, mas narrativa, dada sua estrutura teleológica. A compreensão da vida e de seu *telos* depende, portanto, de sua unidade narrativa. A narrativa não é, dessa forma, um acessório que incorporamos à nossa vida e sim sua estrutura básica. A isso MacIntyre chama de “busca”; e é como “busca” que nossa vida adquire sua constituição narrativa. Quando avaliamos essa “busca”, o que está em questão é a forma que minha vida vai assumir como um todo. Essa questão não pode ser respondida de forma arbitrária, pois o que julgamos não são atos isolados, descontextualizados e com conseqüências de curto prazo. O que está em jogo é o sentido de uma vida como uma totalidade.

O conceito de pessoa, como se depreende do que foi dito acima, pode ser elaborado a partir de duas perspectivas distintas, classicamente consideradas, senão como mutuamente excludentes, ao menos como independentes entre si: a *metafísica* (representada aqui pela tradição lockeana em geral e por Parfit em particular), em que se busca, a partir do ponto de vista de terceira pessoa, produzir uma teoria da identidade quantitativa, ou seja, estabelecer um critério de identidade numérica que elucide o que faz com que sejamos a mesma pessoa através do tempo; e a *prática*, em que se busca prioritariamente produzir uma teoria da identidade qualitativa (sendo a identidade quantitativa um aspecto decorrente desta), na qual o importante é a investigação sobre o tipo de pessoa que sou, através do exame, a partir dos padrões morais da comunidade, dos compromissos e relações pessoais que definem o que seja uma vida plena de sentido. Aqueles que adotam a perspectiva prática (examinaremos aqui especificamente a teoria de MacIntyre) buscam justificar sua posição criticando aqueles que adotam a perspectiva metafísica por produzirem uma teoria que, ao desconsiderar, segundo os defensores da perspectiva prática, o aspecto crucial do conceito de pessoa como um agente moral, acabam por produzir teorias que, por seu caráter reducionista, são incapazes de abarcar todo o sentido não apenas desse conceito, mas também do de ação moral.

Levando-se em conta esse último ponto, é possível descrever uma outra forma de distinguir essas duas perspectivas, partindo-se da análise das

interpretações opostas que ambas adotam quanto ao problema de saber qual é o estatuto ontológico que deve ser atribuído a uma pessoa. Aqueles que adotam a perspectiva metafísica são em geral também reducionistas, pois consideram que uma explicação correta do conceito de pessoa deve necessariamente incluir uma teoria da constituição, que parte de uma instância impessoal ou pré-pessoal, em que os constituintes da pessoalidade (crenças, desejos, ações, etc.) são concatenados através de uma relação R (memória ou *quasi*-memória¹) definida de tal forma que ela seja capaz de preservar a identidade pessoal². Já aqueles que adotam a perspectiva prática geralmente são também anti-reducionistas, adotando, desta forma, uma postura holista ou hermenêutica com relação ao conceito de pessoa, negando a necessidade ou, mais do que isso, a correção de toda abordagem desse conceito que envolva a presunção de uma teoria da constituição que parte de elementos impessoais, vale dizer, da correção de qualquer abordagem reducionista da pessoalidade. Por “hermenêutica” ou “holista” quero dizer uma descrição de pessoa que envolve não apenas a presença de um corpo real e da capacidade reflexiva, que são considerados pelos reducionistas como condições suficientes para a pessoalidade, mas também a presença de *outras pessoas*, admitindo-se, dessa forma, uma certa circularidade na descrição da pessoalidade, circularidade essa que se insere na própria constituição do *self* (se quisermos utilizar o termo “constituição” nesse sentido), que se daria, neste caso, de forma necessariamente intersubjetiva.

Nenhuma dessas explicações do conceito de pessoa pode ser considerada correta, pois, como se verá no decorrer do presente artigo, nenhuma delas cumpre dois requerimentos essenciais, que devem ser observados por qualquer teoria da pessoalidade, quais sejam, primeiro, o de prover um critério fidedigno de distinção entre pessoas reais e ficções³ e, segundo, o de dever poder explicar satisfatoriamente a irracionalidade⁴ (algo especial-

¹ “*Quasi*-memória” é um termo que indica uma representação que se refere a uma certa situação *como se fosse* um evento passado. Essa restrição é necessária justamente para que não seja pressuposta a existência de memórias reais, referente a eventos que realmente ocorreram no passado, evitando-se assim as críticas de que os critérios de identidade pessoal em termos de memória são circulares.

² A alternativa aqui seria ser cartesiano, ou seja, supor ou pretender demonstrar que uma pessoa é uma entidade distinta dos eventos psico-físicos que ela tem; nesse caso, evidentemente, ao invés de constituição, teríamos a tese segundo a qual uma substância pensante é uma entidade simples dada, não redutível a elementos mais simples. Essa alternativa cartesiana é pouco defensável contemporaneamente.

³ A distinção entre pessoas e personagens ficcionais é de senso comum; considerarei essa distinção, no presente artigo como “primitiva”, não requerendo, portanto, posterior justificação.

⁴ Quanto à importância essencial desse último requerimento, faço minhas as palavras de Davidson, em seu artigo seminal, “Paradoxos da irracionalidade”: “Uma teoria que não pudesse explicar a irracionalidade seria uma teoria que também não poderia explicar nossos esforços salutares, e nossos sucessos ocasionais, ao realizarmos a autocritica e o auto-aprimoramento” (Cf. Davidson [1982]).

mente grave para a teoria narrativista de MacIntyre, cuja ênfase recai, exatamente, no aspecto qualitativo, prático da identidade pessoal). Examinaremos neste artigo um aspecto particular da irracionalidade, o fenômeno do auto-engano ou auto-ilusão, fenômeno esse cuja descrição correta exige precisamente, como veremos, a prévia observância do primeiro requerimento, já que a auto-ilusão, como será mostrado, é uma distorção cognitiva que impede a distinção entre narrativas reais e ficcionais, distinção essa que precisa ser bem definida desde o início, sem o que o fenômeno não poderia ser descrito dentro da teoria. A estratégia da argumentação será a seguinte: em primeiro lugar, mostrarei que tanto o narrativismo quanto o modelo neo-lockeano de pessoa se mostram circulares quando se trata de explicar a auto-ilusão, o que as torna teorias incorretas da pessoalidade, já que são incapazes de satisfazer o segundo requisito, e, *a fortiori*, o primeiro requerimento, pois a satisfação do primeiro requerimento está ligada à satisfação do segundo.

O objetivo deste artigo é mostrar que, do ponto de vista meta-ético, as dificuldades da caracterização narrativista do conceito de pessoa e de identidade pessoal levam a tentativa macintyreana de reconstrução da ética das virtudes ao impasse, já que ela exige, como condição de possibilidade para a correta definição do agente moral, um conceito de pessoa capaz de prover um critério bem definido para a identidade pessoal. O exame dessas dificuldades corresponde à parte negativa do artigo; defenderei, na conclusão do artigo, que corresponde à sua parte positiva ou prospectiva, a necessidade de uma reconstrução do narrativismo, escoimado dos vícios acima descritos, para além de MacIntyre, tendo como justificativa que, se partirmos de uma perspectiva pré-filosófica, por razões que descreverei ao final do presente artigo, o narrativismo se mostra possuidor de um alto grau de plausibilidade. No entanto, caso essa reconstrução se mostre inviável no decorrer da pesquisa, isso nos levaria a admitir que a postura correta para o problema da caracterização narrativista do conceito de pessoa, bem como para o problema da caracterização do agente moral a partir da ética das virtudes é a do ceticismo. Tal ceticismo se justificaria devido ao fato de que, por um lado, o conceito de pessoa está (plausivelmente) vinculado à narrativa e, por outro, as tentativas narrativistas para a caracterização deste conceito, ou ao menos aquela apresentada por MacIntyre, fracassaram.

2. Macintyre, Parfit e a auto-ilusão

Examinaremos a seguir dois tipos de concepções de pessoa e de identidade pessoal, uma de persuasão anti-reducionista, a de MacIntyre e o modelo “neo-lockeano” proposto por Parfit, de caráter marcadamente reducionista. O objetivo deste exame será mostrar que nenhuma dessas explicações do conceito de pessoa pode ser considerada como satisfatória, por não satis-

fazerem os dois requerimentos acima descritos. Examinarei em primeiro lugar o modelo reducionista de Parfit, terminando com o exame do narrativismo anti-reducionista de MacIntyre⁵.

2.1 - Apresentação das teses gerais do reducionismo lockeano e neo-lockeano

O reducionismo inerente à maior parte das teorias contemporâneas da identidade pessoal, no sentido de identidade numérica, tal como observamos através do exame dos principais momentos teóricos da tradição inaugurada por Locke e que se segue até os neo-lockeanos, como Parfit ou Perry,⁶ parte da perspectiva metafísica, e se caracteriza pela rejeição das concepções substancialistas do *self*; essas últimas consideram as crenças, os desejos e as ações como modificações de uma substância pensante, que permanece numericamente idêntica a si mesma através do tempo. Para os reducionistas, ao contrário, o conceito de identidade pessoal deve ser caracterizado em termos de uma relação específica entre seus elementos constituintes, isto é, crenças, desejos e ações, sem que seja necessário, dessa forma, fazer qualquer referencia a “fatos adicionais” (como diz Parfit) que envolvam algum tipo de concepção substancialista do *self*, como o *cogito* cartesiano ou a alma cristã. A identidade pessoal é definida em termos relacionais como uma relação R que garante a continuidade psicológica; normalmente, essa relação é caracterizada como a relação de memória, a qual conecta eventos mentais entre si. As teorias da identidade pessoal dos reducionistas neo-lockeanos contemporâneos são uma reconstrução da teoria da memória de Locke, com o objetivo de escoimá-la de suas dificuldades clássicas, como o da circularidade do critério da memória.

Na concepção de identidade pessoal proposta por Locke, a memória é entendida como a relação que preserva a continuidade psicológica, assim definida: no tempo t_2 sou idêntico àquela pessoa em t_1 se e somente se eu posso lembrar as experiências em t_1 como minhas. Ou seja, contra Descartes, Locke nega, não propriamente que *existam* substâncias pensantes que permanecem as mesmas ao longo do tempo, mas sim que tal existência seja uma condição necessária para a atribuição de identidade pessoal e de manutenção da mesmidade numérica de uma só pessoa ao longo do tempo:

Essa “intuição lockeana” é ao mesmo tempo extremamente simples e bastante difícil de ser refutada. Ela parece mostrar de forma inequívoca que

⁵ A perspectivista narrativista não está relacionada necessariamente com uma postura anti-reducionista com relação ao *self*; há de fato uma variedade eliminativista de narrativismo representada por Dennett [1991] e [1992]. Para uma crítica ao projeto dennettiano de naturalização da consciência, cf. Fonseca [2004].

⁶ Cf. Perry [1975] e Parfit [1984].

a identidade de uma mesma substância mental não é uma condição necessária nem suficiente da identidade pessoal (nem da numérica nem da qualitativa; restringimo-nos por ora à primeira). Várias substâncias podem se suceder, mas, desde que o fluxo da consciência, conectado pela memória, seja mantido, podemos afirmar que se trata de uma mesma pessoa; a mesma substância pode permanecer, mas, sem a memória dos eventos passados que lhe aconteceram, ela não pode ser dita a mesma pessoa. Por outro lado, a continuidade do mesmo corpo tampouco parece ser necessária nem suficiente para a manutenção da identidade pessoal, como mostra o exemplo do príncipe e do mendigo⁷: se a alma do príncipe migrasse para o corpo de um mendigo, preservando consigo todas as conexões de memória do príncipe, ninguém, diz Locke, duvidaria de que aquele indivíduo corporal é responsável pelas ações do príncipe, e não do mendigo – ou seja, em um sentido moral e jurídico, mas que revela nossos pressupostos ontológicos, a pessoa que tem o corpo do mendigo é na verdade o príncipe.

Como mostrou um dos primeiros críticos dessa teoria, Butler⁸, o critério da memória para a identidade pessoal proposto por Locke é falso, devido ao seu caráter circular. A crítica da circularidade formula-se mais ou menos nestes termos (não sigo aqui a formulação de Butler, mas uma forma geral que poderia assumir a crítica): a identidade pessoal é definida por Locke como sendo constituída através de relações de memória; assim sendo, é preciso que haja um critério suplementar para distinguir pessoas falsas de pessoas verdadeiras, ou de forma mais intuitiva, um critério de distinção entre memórias falsas e memórias verdadeiras, sendo as memórias falsas aquelas produzidas por algum tipo de relação causal externa (como um hipnotizador) ou mesmo interna (como ilusões de todo o tipo)⁹. Mas o único critério possível de distinção entre memórias verdadeiras e memórias falsas, considerando que a identidade pessoal é definida a partir das primeiras, teria de ser formulado em termos da identidade da *pessoa* que possui as lembranças reais, pois, a partir dessa perspectiva de primeira pessoa, levando em conta apenas os conteúdos apresentados pela memória, essa é a única maneira pela qual poderei dizer que, de fato, não sou aquela pessoa constituída pelas relações de falsas memórias, ou seja, que não me identifico com essa pessoa ficcional constituída pelas conexões de falsas memórias. Uma memória é real se ela recupera uma experiência que a *mesma* pessoa que tem a memória teve realmente no passado; mas para enunciar essa condição, tivemos de lançar mão da identidade pessoal, que

⁷ Locke [1979] § 15.

⁸ In: Perry [1975], p. 99-106.

⁹ É interessante notar que quando se introduz nesta discussão fenômenos como a auto-ilusão, mesmo quando consideramos as reformulações neo-lockeanas do critério da memória, o problema da circularidade reaparecerá, porque em ambos os casos é produzido algo como memórias falsas, por uma decisão interna e prospectiva, como será mostrado mais adiante.

era o que presumivelmente estava sendo definido pela relação de memória. Não há, assim, concluir a crítica da circularidade inspirada em Butler, um critério interno de distinção entre pessoas reais e ficcionais (e ele terá de ser interno, devido ao próprio caráter “fenomenológico” com o qual Locke define o que seja uma pessoa real – isto é, a partir dos conteúdos da memória apreendidos por um sujeito pensante) que não seja circular.

O critério da memória proposto por Locke foi reformulado contemporaneamente por filósofos que chamarei de “neo-lockeanos”, cujo expoente maior é Derek Parfit, com o objetivo de eliminar o problema da circularidade, inerente à perspectiva “fenomenológica” a partir da qual esse critério foi originalmente concebido. Essa reconstrução consiste no abandono da perspectiva “fenomenológica” da teoria de Locke, redefinindo a memória em termos de relações causais, descritas impessoalmente, ou seja, em termos de terceira pessoa, com a vantagem de que agora é possível referir-se a uma pessoa como idêntica a si mesma de forma não circular.

Parfit ([1984] p. 199 e ss.) considera que pessoas são constituídas por eventos mentais e físicos, não havendo necessidade de se levar em conta “fatos adicionais”, tais como a existência da alma ou do ego cartesiano, que permaneceriam idênticos a si mesmos através de suas mudanças internas (crenças, desejos, afetos, etc). A partir dessa visão reducionista com respeito à pessoalidade, a identidade pessoal através do tempo é entendida por Parfit como consistindo em uma relação particular entre os elementos constituintes de uma pessoa, i. e., uma inter-relação entre eventos mentais e físicos que ocorrem em uma região bem determinada do espaço-tempo, qual seja, um corpo humano, ou, ao menos, um cérebro humano.

Advogando o critério psicológico (superveniente a eventos físicos do cérebro) como o único capaz de descrever corretamente o conceito de identidade pessoal, Parfit define essa última da seguinte forma: *A* e *B* são a mesma pessoa se e somente se há entre elas relações psicológicas, denominadas “relações *R*”, sendo que essas relações devem possuir o tipo apropriado de causas, ou seja, aquelas que preservam a continuidade espaço-temporal do corpo (ou, ao menos, do cérebro) e as relações psicológicas que lhe são correspondentes, e não devem, além disso, produzir séries ramificadas, isto é, não deve haver uma outra pessoa *C*, cujas relações psicológicas sejam continuções de *A*. As relações *R* são de dois tipos, as conexões psicológicas e a continuidade psicológica. Conexões psicológicas entre *A* e *B* são relações psicológicas diretas, cujos conteúdos podem ser conexões de memória, de ações, de caráter ou de crenças e desejos; se as conexões psicológicas diretas entre *A* e *B* se estruturarem de tal forma que, diariamente, ao menos a metade delas permaneçam, então elas compõem o que Parfit chama de conexões psicológicas fortes. A continuidade psicológica entre *A* e *B* é definida como um emaranhado de conexões psicológicas fortes, tal como os diversos fios que compõem uma corda; com isso,

o critério psicológico de identidade pessoal proposto por Parfit possui a seguinte forma: Sejam *A* e *B* pessoas. *A* é idêntica a *B* se e somente se há uma relação de continuidade psicológica não bifurcada entre *A* e *B*¹⁰.

2.2 - A auto-ilusão e a circularidade do narrativismo de MacIntyre

Uma vez apresentado esse resumo da teoria neo-lockeana da constituição do *self* e da identidade pessoal, podemos agora voltar à teoria narrativista de MacIntyre e nos perguntarmos como ela se relaciona com a teoria neo-lockeana. Isso nos ajudará a esclarecer quais são os problemas teóricos presentes nesses dois tipos opostos de teoria.

No parágrafo anterior à sua crítica ao reducionismo lockeano e neo-lockeano (que, em suma, busca mostrar que as dificuldades encontradas por Locke e seus sucessores se devem ao caráter reducionista de suas teses sobre a pessoalidade¹¹), MacIntyre introduz o problema da constituição do *self*, embora não em termos explícitos:

Uma tese central começa, então, a emergir: o homem em suas ações e prática, tanto quanto em suas ficções, é essencialmente um animal contador de histórias. *Ele não o é essencialmente, mas torna-se, através de sua história*, um contador de histórias que aspiram à verdade (MacIntyre [1985] p. 216; grifo meu).

Ele esboça assim, através da restrição expressa na frase grifada por mim nessa citação, a necessidade de uma teoria narrativista da constituição de uma pessoa: nós nos incorporamos à sociedade humana através da incorporação de um ou mais personagens. Temos de aprender o que é ser esse (s) personagem (ens) para que possamos compreender nossas responsabilidades e as responsabilidades dos outros e para que nos tornemos as pessoas que emergem dessas narrativas. As crianças aprendem isso através de histórias que lhes são contadas e de histórias sobre si mesmas que elas aprendem paulatinamente a contar (para si mesmas e para os outros).

No entanto, não fica claro como algo que ainda não é uma pessoa, i. e., um organismo humano vivo, se torna um contador de sua própria história. O esclarecimento desse processo de constituição de uma pessoa implicaria a adoção, ao menos neste âmbito restrito da explicação do que seja a pessoalidade, da perspectiva metafísica, ou seja (uma vez que MacIntyre definitivamente não é um cartesiano), reducionista, como a de Locke e dos neo-lockeanos. Mas MacIntyre toma exatamente o caminho inverso; ele passa imediatamente, no parágrafo seguinte, a atacar a perspectiva

¹⁰ Nesse caso, devemos entender *A* e *B* como estágios temporais de pessoas.

¹¹ MacIntyre [1985], p. 216 e seq.

metafísica. Ele parte da tese de que as dificuldades da perspectiva lockeana e neo-lockeana, representada por Parfit, se devem ao fato de que elas não levam em conta o *background* narrativo em sua descrição da identidade pessoal. A dificuldade principal a que MacIntyre parece se referir é a da circularidade. No caso de Locke, essa dificuldade, como vimos mais acima, é bem conhecida; já no caso dos neo-lockeanos, há várias alternativas para evitar a circularidade do critério da memória (a principal estratégia, de propor uma teoria causal da memória, também já foi exposta em linhas gerais mais acima). Para MacIntyre, a dificuldade parece se resolver metodologicamente, através da adoção de uma visão holística ou hermenêutica do processo de constituição do *self*: a identidade pressupõe a unidade de um personagem; a unidade do personagem, por seu turno, pressupõe sua identidade. A responsabilidade moral pressupõe ambas.¹² Segundo esse método, que designamos aqui como “hermenêutico”, é preciso evitar uma visão reducionista do *self*, e a única maneira de se fazer isso é supor desde o início uma perspectiva descritiva em primeira pessoa, segundo a qual uma pessoa é constituída por relações narrativas em uma comunidade dada, na qual outras pessoas introduzem-na nas práticas sociais formadoras da personalidade. Uma explicação da pessoalidade já supõe, assim, de antemão, uma referência a outras pessoas, de forma ineliminável. O círculo lógico é assim substituído pelo círculo hermenêutico.

Mesmo que aceitemos, com MacIntyre e os anti-reducionistas, essa superação metodológica do problema da circularidade da abordagem neo-lockeana da pessoalidade, através da aceitação de uma espécie de “círculo virtuoso” que se confundiria com o próprio método filosófico adotado por eles, essa abordagem holística não pode, em seus próprios termos, dar conta do conceito de pessoa, devido ao fato de que tal tipo de narrativismo não satisfaz o segundo requerimento para que uma teoria da pessoalidade seja considerada correta, que é o de dever poder explicar a irracionalidade, explicação sem a qual não é possível caracterizar satisfatoriamente uma pessoa como um agente moral. Como mostraremos mais adiante, o narrativismo de MacIntyre, ao não ser capaz de explicar o fenômeno da auto-ilusão, e, por isso mesmo, não ser também capaz de satisfazer ao primeiro requerimento, o de dever poder exibir um critério fidedigno de distinção entre pessoas e ficções (ou: personagens reais e ficcionais), não pode ser considerado como uma explicação correta do conceito de pessoa (Lembremos que a auto-ilusão é definida exatamente como um desvio cognitivo cuja característica distintiva é impedir a distinção entre narrativas reais e ficcionais, o que pressupõe que a teoria, qualquer que seja ela, seja internamente capaz de representar, de antemão, esse critério de distinção, que agora sucumbe, com a instauração do fenômeno da auto-ilusão,

¹² MacIntyre [1985] p. 218.

ou em outras palavras: a incapacidade de explicação da auto-ilusão revela exatamente a incapacidade da teoria de prover um critério de distinção entre realidade e ficção).

Tratarei agora de mostrar que, ao apresentar um modelo de constituição narrativa do *self* (segundo o qual as ações e estados mentais de um organismo humano devem ser conectados pelo que chamarei aqui de relação N (de narrativa) para que se constitua a partir disso uma pessoa), o narrativismo se mostra inconsistente. A estratégia para desenvolver essa crítica será, como disse acima, examinar um fenômeno estreitamente conectado à identidade pessoal qualitativa, indiretamente conectado à identidade numérica e muito importante para nossa vida moral, a saber: o fenômeno da auto-ilusão. A importância da compreensão desse fenômeno para uma teoria moral mostra-se diretamente pela consideração de que dificilmente poderíamos, sem ele, explicar a escolha do mal, pois normalmente acreditamos estar fazendo o bem mesmo em nossos piores pecados. Sendo assim, toda teoria, como a narrativista, que tivesse pretensões de desenvolver uma explicação sobre a dimensão moral de nossa personalidade teria de poder explicar, senão a irracionalidade em geral, ao menos esse fenômeno em particular. Mostrarei, porém, que o narrativismo não consegue apresentar um modelo teórico consistente para explicar a auto-ilusão.

Compreendemos bem a estrutura da mentira: mentir é causar uma crença falsa em outra pessoa. Mas o que significa dizer que alguém engana a si próprio? Isso implicaria a situação em que crenças falsas e crenças verdadeiras conviveriam sem aparentemente produzir qualquer embaraço cognitivo ao sujeito que as possui, apesar de ele saber discriminar tais crenças como falsas e verdadeiras. Do ponto de vista formal, a auto-ilusão pode ser definida da seguinte maneira: Sejam S e S' sujeitos. Diz-se que S se auto-ilude se S crê que p , e S causa em S' a crença de que $\sim p$, sendo $S = S'$. Se S é diferente de S' , diz-se apenas que S engana S' . Visto que no conceito de enganar está a incluída intenção de produzir uma mentira, o conceito de auto-engano possui a característica contraditória de que S terá de crer simultaneamente em p e $\sim p$. Como um fenômeno real não pode ser contraditório, isso parece indicar que a auto-ilusão não é um fenômeno real (Poderíamos também suspeitar que há um erro inicial em se caracterizar a auto-ilusão a partir do modelo da mentira, já que esse último envolve um elemento intersubjetivo ausente do primeiro; em certo sentido, como veremos, isso é verdade, mas, ainda assim, será salvaguardado um sentido comum a ambos, que nos permitirá afirmar que eles são dois aspectos distintos da mesma atividade básica de mentir, a si mesmo ou a outrem).

Devido a essas dificuldades, caberá ao cético com respeito ao fenômeno da auto-ilusão mostrar que o que de fato existe são fenômenos bastante similares, cuja descrição não possui características contraditórias. Há duas formas possíveis e opostas para descrever tal fenômeno. Em primeiro lugar,

o cético poderia propor que o que existe de fato é um erro, uma apresentação incorreta de minhas próprias representações. Esse engano é intencional, mas apenas no sentido estrito de ser uma experiência consciente, e não no sentido de ser produzido propositalmente – ninguém *quer* errar nessa acepção do termo “intencional”. No entanto, essa descrição cética da auto-ilusão extrai seu sucesso da trivialização do fenômeno, retirando-lhe seu aspecto moral. Contra a solução cética, parece ser verdade que a auto-ilusão é uma mentira que contamos a nós mesmos, e como toda mentira, é objeto de reprovação moral. Mas a descrição proposta pelo cético reduz o fenômeno a um mero erro cognitivo sobre nós mesmos, o que retira o aspecto proposital da auto-ilusão. Outra proposta consideraria que o que de fato existe é um mecanismo não intencional, no qual razões são meras causas, ou seja, onde nossas inclinações teriam um poder editorial sobre nossa apreensão dos fatos, fazendo com que os aspectos indesejáveis sejam sub-repticiamente deixados de lado, em benefício dos aspectos favoráveis. Essa proposta também é cética, porque agora é o aspecto intencional da auto-ilusão que é deixado de lado, em qualquer um dos dois sentidos de “intencional” discriminados acima, pois, por essa descrição, *algo* me engana, e não eu mesmo (ainda que esse “algo” esteja no interior de minha mente), já que as razões para o engano funcionam como meras causas, o que não permite tampouco a imputação moral.

O que se pode aprender dessas duas descrições céticas é que a descrição correta da auto-ilusão, se é que ela existe, necessita de uma abordagem que compatibilize as duas propostas, pois a compreensão comum deste fenômeno parece envolver tanto aspectos intencionais quanto causais, formando um amálgama que abrange tanto o saber de si quanto a perda de controle sobre si. Essa perda de controle sobre si se dá, como mostrarei a seguir, devido a uma distorção cognitiva em que o ficcional e o real se tornam indiscerníveis. A formulação de um modelo para explicar esse fenômeno mostrará justamente que ele *pode* existir. A segunda consequência da elaboração desse modelo será a prova da falsidade do narrativismo.

Mas, é plausível não ser cético com relação à auto-ilusão? Há algo em nossa experiência cotidiana que indique essa plausibilidade? Para esclarecer o fenômeno da auto-ilusão e mostrar que sua existência é defensável, adotarei a estratégia de compará-lo com um fenômeno que, embora compartilhe de estruturas similares, desfruta da vantagem de não despertar suspeitas quanto ao seu estatuto ontológico e sua realidade. Usarei o exemplo da fruição estética. Antes de continuar, porém, farei brevemente uma descrição das relações entre crenças e narrativas no modelo narrativista, para, logo após, examinar a relação particular entre crenças e narrativas no caso específico do fenômeno da auto-ilusão.

Um critério possível de distinção entre narrativas ficcionais e reais poderia se basear na relação dessas últimas com nossas crenças. Para esclarecer

esse ponto, é necessário explicar, mesmo que seja de forma preliminar, o que constitui uma obra literária enquanto tal. Quando lemos um romance (ou vemos um filme), o qual passa a nos interessar, somos, por assim dizer, como que “tragados” por seu campo de gravidade ficcional, fazendo-nos vivenciar, a despeito de sua patente incongruência com os fatos que se sucedem no mundo real, os acontecimentos do mundo ficcional como se fossem reais, ocorrendo assim aquilo que Coleridge denominou de “suspensão da descrença”, algo como a operação fenomenológica, em todo o caso espontânea e não devido aos rigores de caráter metodológico, de “pôr entre parênteses” a realidade. Voltando ao nosso problema, podemos dizer agora de que modo podemos demarcar o limite entre as narrativas reais e as narrativas ficcionais. As narrativas ficcionais têm o que podemos chamar de “relação de segunda ordem” com nossas crenças, que consiste no fato de que elas são constituídas pela experiência estética da suspensão da descrença, enquanto que as narrativas reais teriam a “relação de primeira ordem” de *constituir* nossas crenças. Como as relações de segunda ordem são hierarquicamente dependentes das relações de primeira ordem, não precisamos recorrer à existência prévia de pessoas reais como critério de distinção entre biografias e romances. Assim sendo, mesmo que, como querem os narrativistas, pessoas reais não pré-existam às narrativas, mas sejam de fato constituídas por elas, há, no entanto, um critério preciso (formal, isto é, interno à forma narrativa, à função de narrar e à descrição das crenças do narrador) de distinção entre narrativas reais e ficcionais e entre personagens reais e ficcionais. Esse critério se revelará, no entanto, insatisfatório, devido ao fato de que a auto-ilusão se desenvolve exatamente no nível das relações de primeira ordem, justamente aquelas que nos constituem como pessoas, pois, como se mostrará, é exatamente por interferir no próprio processo de constituição do *self* que a auto-ilusão não poderá ser explicada dentro da teoria, sob risco de circularidade.

Voltemos à “suspensão da descrença”. Deixaremos momentaneamente de lado o contexto original de sua descrição, para nos concentrarmos apenas na comparação de sua estrutura com a da auto-ilusão. Tal como no caso da auto-ilusão, também no da suspensão da descrença opera-se um “desvio cognitivo”, onde o ficcional ganha, mesmo que de forma intermitente, foros de realidade¹³, como no caso da leitura que absorve nossa atenção. Podemos dizer que a fruição estética é fruto de uma intenção anterior de se operar um “relaxamento” em nossa capacidade discriminatória com relação à distinção entre o real e o ficcional? Ao menos em parte isso parece ser verdade. De fato, podemos nos dispor à fruição estética, mas não podemos produzi-la a partir de um ato de decisão. Ao menos alguns

¹³ Sartre, em *O ser e o nada* (Capítulo 2, “A má-fé”, Seção I, “Má-fé e mentira”; [1943] [1997]) mostra que a má-fé (que designa, em seu vocabulário, o que chamamos aqui de ‘auto-ilusão’) tem uma estrutura “meta-estável”.

de nossos atos de fruição estética são de fato cercados por algumas cerimônias preparatórias, as quais, segundo nossa forma de socialização, nos propiciam exatamente as condições para tal “relaxamento”, como o desligar das luzes antes do começo de uma peça. Assim sendo, ao que parece, uma condição necessária para que ocorra o fenômeno da suspensão da descrença é uma decisão anterior que nos faz estar predispostos à sua ocorrência, mas, no entanto, essa decisão não é uma condição suficiente, pois a suspensão da descrença pode simplesmente não ocorrer, como, por exemplo, no caso de uma peça ruim.

Do ponto de vista de primeira pessoa, a fruição estética aparece como uma descontinuidade, ou como uma “transição de fase”, que ocorre em dadas condições, sem que possamos controlá-las ou prevê-las¹⁴, mas apenas nos dispor para a sua ocorrência. Assim sendo, o caráter não-intencional ou incontrolável dessa transição de fase deve ser descrito em termos de causas, e não de razões. Podemos explicar que a identificação de alguém com um personagem ficcional se deve ao fato de que a trajetória narrativa de tal personagem coincide com vários traços importantes da própria narrativa real dessa pessoa. Neste caso, razões funcionam como causas: é o desejo de auto-compreensão o aspecto motivacional que propiciou o relaxamento cognitivo descrito como a identificação com o personagem, que só pode ocorrer devido à suspensão da descrença.

A estrutura do fenômeno que nos interessa, a auto-ilusão, guarda similaridades com esse último exemplo. No caso da fruição estética, não estamos mentindo para nós mesmos, pois a identificação com o personagem se dá em um contexto meramente ficcional, ou dito de outra maneira: identificar-se com o personagem não é mentir para si mesmo, pois onde não há pretensões de verdade não pode haver a intenção de mentir tanto para outros quanto para nós mesmos. Podemos agora caracterizar a auto-ilusão por comparação com o fenômeno da fruição estética, de tal forma a eliminar as versões céticas do fenômeno que pretendiam reduzi-lo a fenômenos diferentes, nos quais o modelo da mentira interpessoal não se aplicaria de forma alguma. A auto-ilusão é constituída por três instâncias: razões se tornam causas, há uma “perda de controle sobre si” e também há a intenção de mentir, que lhe dá uma dimensão de culpabilidade, tal como comumente ela é compreendida. Temos agora os meios para definir a auto-

¹⁴ Não faz parte do escopo deste artigo a produção de uma teoria da fruição estética; para efeito de argumentação, basta levar em conta o aspecto descontínuo desse fenômeno tal como ele aparece à perspectiva de primeira pessoa. Como a perspectiva de primeira pessoa é fundamental para a tese narrativista, tal teoria não desempenharia nenhum papel relevante para a minha argumentação a seguir, que tratará de mostrar que o narrativismo é uma teoria errada da constituição da identidade pessoal, levando em conta apenas o modo como a experiência da auto-ilusão pode ser descrita em primeira pessoa.

ilusão¹⁵. Trata-se de uma disfunção cognitiva na qual, devido a aspectos motivacionais (desejos), eu intencionalmente me disponho a estar em uma situação (futura e previsível) onde tais desejos funcionarão como causas para que não mais se possa distinguir entre realidade e ficção, ou, de outra forma, me disponho a me colocar intencionalmente em uma situação futura onde tais crenças me causarão a crença que $\sim p$, mesmo que agora seja verdade que eu saiba que p ¹⁶.

Dada essa definição de auto-ilusão, vejamos agora por que a teoria narrativista não pode satisfazer aos critérios que permitiriam distinguir narrativas reais de auto-ilusões, sem o que ela não seria um modelo satisfatório desse fenômeno. A reformulação narrativista do conceito de auto-ilusão terá de ser construída em termos das relações entre o *self* e suas crenças de primeira ordem, narrativamente constituídas; em outras palavras, é o indivíduo enquanto personagem que se auto-ilude¹⁷, já que uma pessoa, para o narrativista, é caracterizada, pela relação N, como um personagem real. A auto-ilusão seria uma fração de uma auto-narrativa, e assim, intersubjetivamente contextualizada. Antecipemos desde já que não é possível tal reformulação, e que essa impossibilidade se deve a uma limitação expressiva inerente à caracterização narrativista do *self*, como mostraremos a seguir. Isso equivale a dizer que o narrativismo é de fato uma teoria insuficiente com relação ao *self*, já que há aspectos importantes de nossa razão prática, como mostraremos a partir do exemplo da auto-ilusão, que deixam de ser explicados.

¹⁵ Adoto aqui a definição de auto-ilusão encontrada em Pinheiro [manuscrito não publicado]. Essa definição pode ser considerada como uma modificação da definição de auto-ilusão proposta por um dos teóricos mais influentes deste campo de estudos, que é Mele [1997]. É claro que precisaremos, no decorrer da pesquisa, investigar algumas das mais importantes definições contemporâneas de auto-ilusão, que podemos considerar como clássicas desse campo, que são as de Davidson [1982], que supõe que a explicação da irracionalidade em geral, e da auto-ilusão em particular, precisa supor a divisão funcional da mente (algo que pode encontrar ecos em uma futura reconstrução do narrativismo, já que a divisão funcional do *self* já é pressuposta no narrativismo de MacIntyre, ou seja, a divisão funcional entre autor e personagem) e a de Audi [1982], que supõe, para explicar a irracionalidade, não a divisão da mente, e sim a existência de pensamentos inconscientes, e sua intervenção sobre os pensamentos conscientes, à la Freud (o que também não precisa ser descartado em uma futura reconstrução do narrativismo, talvez através de uma relação específica entre a função de autor e a função de personagem).

¹⁶ Exemplos triviais de auto-ilusão seriam: penso que meu filho, que apresenta todos os sinais externos de um viciado em drogas, é na verdade um estudante exemplar, graças a um processo de ignorar e interpretar evidências em favor dessa crença (auto-ilusão a respeito de outros); penso, contra todos os sinais aparentes, que na verdade minha doença não é fatal, e que estou curado, apenas ainda um pouco abatido pelo período difícil por que passei (auto-ilusão a respeito de mim mesmo).

¹⁷ Resta saber, se aceitamos que de fato é o personagem que se auto-ilude, qual seria a função do autor no processo de auto-ilusão. Cf. as sugestões dadas na nota 15.

Comecemos nosso argumento contra o narrativismo apontando para uma sua dificuldade inicial. O que poderia significar descrever um personagem como enganando-se a si mesmo no interior de uma narrativa? No caso do narrativismo, a auto-ilusão tem de se dar ao nível das relações de primeira ordem entre crenças e narrativas, pois é preciso explicar como uma certa pessoa conta para si mesma a sua vida de tal modo a incorporar, entre as suas crenças, a crença de que *p* (por exemplo, a crença de que seu filho não é drogado)¹⁸.

No entanto, faz parte da estrutura da auto-ilusão a suspensão da descrença que, aparentemente, só faz sentido no nível das relações de segunda ordem entre crenças e narrativas, pois é preciso entendê-la como uma certa relação do sujeito com algumas de suas crenças – a suspensão da descrença com relação a certas evidências frágeis que “sustentam” certas crenças que satisfazem nossos desejos (por exemplo, o desejo de que meu filho não seja drogado). A auto-ilusão, reformulada em termos narrativistas, tem como característica básica de sua estrutura a suspensão da descrença com relação a uma história falsa que contamos para nós mesmos, enquanto personagens de uma história real. Isso supõe que nós, de alguma forma, discriminamos as narrativas reais das narrativas ficcionais, e que, no entanto, como resultado da atuação de aspectos motivacionais (como p. ex., desejos), dissimuladamente desprezamos tal distinção¹⁹.

Podemos agora formular diretamente nossa crítica ao narrativismo. A dificuldade insuperável que envolve a descrição narrativista da auto-ilusão é a de que não é possível expressar o duplo aspecto, isto é, o intencional e o causal, que os fatos psicológicos assumem para a estruturação do fenômeno, pois ela se restringe, de forma constitutiva, ao aspecto intencional. Para mostrar isso, é necessário esclarecer as relações entre narrativas e

¹⁸ A constituição de primeira ordem das crenças de um personagem real através das narrativas reais não garante a correta identificação do personagem real com o que podemos chamar de seu papel real. A loucura é o exemplo mais contundente disso. Dessa forma, a identidade pessoal narrativa não pode ser o resultado de algo como uma introspecção em primeira pessoa, e isso quer dizer que podemos nos enganar sobre nós mesmos, isto é, nos enganar sobre o personagem com o qual deveríamos nos identificar, e sobre os papéis que deveríamos representar. A constituição da auto-identidade não é uma tarefa que pode ser creditada apenas, e isoladamente, ao próprio indivíduo, sob pena de auto-ilusão. Essa tarefa cabe também a terceiros realizar, ou melhor, ela só pode ser levada a cabo através da inserção do indivíduo nas relações sociais, e isso se dá de forma constitutiva. Há assim uma dimensão pública de nossa identidade pessoal.

¹⁹ Não abordarei neste artigo a questão de saber em que consiste essa “dissimulação”. Ao que parece, a dissimulação que constitui o auto-engano teria de ser, de alguma forma, relacionada a alguma instância “inconsciente” ou “pré-consciente” do *self*, pois não é o caso que alguém deliberada e expressamente planeje se auto-enganar. Seria preciso, pois, determinar em que consiste descrever narrativamente tal coisa. Talvez possamos combinar a referência a pensamentos inconscientes para explicar tanto a auto-ilusão quanto a constituição narrativa do *self*, a partir talvez de uma modificação da definição de auto-ilusão proposta por Audi, já mencionado na nota 15.

causas. Em primeiro lugar, as narrativas podem conter, sem qualquer embaraço, descrições de processos causais (i. e., o que é narrado pode ser uma relação causal); em segundo lugar, narrativas podem ser causas (de outras narrativas, de estados mentais ou de estados físicos). Por exemplo, é possível descrever narrativamente, ou seja, em termos das relações entre personagens reais, que fui eu que causei o medo, com minha história de terror, que faz com que aquela criança não consiga dormir, ou que a falsa crença de alguém foi causada por minha mentira.

A descrição narrativista da auto-ilusão não pode, no entanto, basear-se no fato trivial de que narrativas causam narrativas, muito embora isso seja verdade *do ponto de vista de terceira pessoa*. Ou seja, é claro que alguém pode narrar meu auto-engano em termos de relações causais entre partes de minhas auto-narrativas, mas isso não acrescenta nada para mim, aquele que se auto-engana, pois para mim o que se dá é apenas um fenômeno de descontinuidade narrativa, que é a suspensão de descrença. A descrição narrativista da auto-ilusão tem como objetivo explicar narrativamente como personagens reais podem suspender a descrença com relação a narrativas reais que os constituem (isto é, que constituem suas crenças). No entanto, é justamente isso que se dá através de um processo que não pode ser expresso para mim mesmo, ou seja, no interior de minha auto-narrativa.

Ora, o elemento causal faz parte de nossa compreensão do fenômeno do auto-engano, como vimos acima (Deve-se notar que afirmar que há um elemento não intencional na gênese da auto-ilusão não é equivalente a afirmar que a causa de minha auto-ilusão são processos inconscientes e, por isso, exteriores ao *self*. Tal caracterização da auto-ilusão é insuficiente, como vimos anteriormente, pois não é capaz de captar o que mais importa nesse fenômeno, que é seu aspecto moral. Segundo essa caracterização, não posso ser responsabilizado por minhas crenças serem ilusórias, já que não sou eu, e sim processos mecânicos além de meu controle, que causam meu desvio cognitivo. No entanto, a auto-ilusão é um fenômeno irreduzivelmente intencional, pois *sou eu* aquele *me* engana, e não processos mecânicos. Ao lado do elemento causal, deve figurar também um elemento intencional).

Logo, da mesma forma que no caso da fruição estética, também a auto-ilusão se estrutura necessariamente em aspectos intencionais e causais: para ser enganado por mim mesmo, é necessário que eu me disponha a um relaxamento cognitivo que, por seu turno, sei de antemão que me levará a um estado onde determinados aspectos motivacionais deixarão de funcionar como razões para se tornarem causas de minha falsa crença futura. Ou, em outras palavras: sou *eu* o responsável pelo relaxamento cognitivo que *sei* de antemão que *me* levará a crer que $\sim p$, isto é, que *me causará* a crença que $\sim p$, sendo que, neste momento, sou *sabedor* que é verdade que *p*.

É uma verdade trivial que a auto-ilusão não pode ser descrita na perspectiva de primeira pessoa, sob pena de contradição, devido ao duplo aspecto,

causal e intencional, dessa relação do *self* consigo mesmo. Ou seja, quem se auto-ilude não pode decidir *reflexivamente* se auto-iludir. O que torna essa conclusão aparentemente trivial importante é o fato de que, sendo o narrativismo constitutivamente uma teoria em primeira pessoa, sua capacidade expressiva só poderia dar conta do aspecto intencional da auto-ilusão, o que torna impossível a correta descrição desse fenômeno. Vejamos o que isso significa. Como uma teoria da constituição do *self*, o narrativismo deve estabelecer quais são as condições necessárias para a identidade pessoal, entendida em termos de unidade narrativa. Do ponto de vista formal, a teoria narrativista define a identidade pessoal nos seguintes termos: seja *S* uma seqüência de auto-narrativas. A identidade pessoal é uma relação *N* entre os elementos dessa seqüência, que garante a unidade narrativa de *S*. Dessa forma, *N* é uma condição necessária, mas não suficiente para a identidade pessoal. Isso não é problema para o narrativista, que não veria qualquer embaraço em concordar que a conceituação correta de identidade pessoal precisa levar em conta outros fatores além da unidade narrativa, como, p. ex., a continuidade física de um corpo humano. Ou seja, o narrativismo poderia admitir que há relações causais (não-intencionais) que são parte das condições necessárias da constituição de pessoas.

Para explicar por que o caráter de primeira pessoa representa um limite interno da teoria narrativista da identidade pessoal, usarei a estratégia de compará-la com uma teoria rival, a teoria neo-lockeana, para mostrar que, por ela ser constitutivamente uma teoria em terceira pessoa, isso a torna capaz de descrever adequadamente, isto é, levando em conta tanto o aspecto intencional quanto o causal, a auto-ilusão, explicando-a como uma relação do *self* com ele mesmo que preserva a identidade pessoal. A teoria neo-lockeana descreve a identidade pessoal em termos de continuidade psicológica entre estágios temporais de pessoas. Diz-se que uma série de estágios temporais de pessoas é psicologicamente contínua se há entre esses estágios uma relação psicológica *R* que preserva a identidade pessoal. Nesse contexto teórico, a auto-ilusão pode ser descrita como uma relação entre estágios de pessoas que preserva a continuidade psicológica, onde, em determinado estágio *E*, dá-se a decisão de pôr-se intencionalmente em um estado cognitivo no qual, em uma situação futura, os desejos que agora aparecem ao sujeito como razões funcionarão como causas de sua crença de que $\sim p$, mesmo que agora seja verdade que ele saiba (ou creia saber) que *p*. Descrita dessa forma, a auto-ilusão se torna um caso especial da relação *R* entre estágios de pessoas; assim sendo, ela é uma relação do *self* consigo mesmo que preserva a continuidade psíquica.

O que o narrativista não pode admitir, visto que ele caracteriza a identidade pessoal em termos intencionais, ou seja, descrita em termos de uma relação necessária *N* entre auto-narrativas, é que haja uma relação do *self* consigo próprio que envolva (ou seja redutível a) relações causais. Que haja relações causais, além das relações narrativas, que contribuam para a

constituição de uma pessoa, o narrativista, como vimos, não se importaria em concordar, já que, segundo ele, as relações narrativas são condições necessárias, mas não suficientes, para a identidade pessoal. O que o narrativista não poderia aceitar é que haja conexões causais mecânicas cegas, não-narrativas, entre dois estados intencionais de um indivíduo (acreditar que p em t_1 e em $\sim p$ em t_2), relações essas que o constituiriam como uma mesma pessoa, pois admitir isso é reconhecer que há uma relação não-narrativa conectando os estados intencionais de uma pessoa, o que é contra as teses mínimas do narrativismo, tal como ele está descrito mais acima. No entanto, como mostramos com a descrição neo-lockeana da auto-ilusão, há uma relação do *self* consigo próprio que envolve não só relações narrativas, mas também relações causais, e que, a despeito disso, preserva a identidade pessoal. Concluimos, portanto, que a unidade narrativa, tal como ela é descrita pelo narrativista, não é uma relação necessária para a identidade pessoal, o que torna o narrativismo uma teoria incorreta da constituição do *self*.

Chegamos deste modo à surpreendente conclusão de que o narrativismo é uma teoria falsa do *self*, pois aspectos fundamentais da vida moral não podem ser expressos na teoria. Esse fracasso reveste-se de especial importância, devido ao fato de que o narrativismo, sendo incapaz de explicar a auto-ilusão, vê-se impossibilitado de realizar sua tarefa mais simples, que é a de oferecer critérios fidedignos de distinção entre personagens reais e ficcionais, já que ele não é capaz de prover um critério de distinção entre auto-narrativas verdadeiras e auto-ilusão. Ou seja, porque a teoria narrativista não é capaz de integrar em seu interior uma explicação de como relações causais entre estágios temporais de pessoas *produzem* narrativas falsas (auto-ilusórias), ela não pode tampouco explicar como essas auto-imagens narrativas falsas surgem nem o que as distingue das eventuais auto-imagens narrativas verdadeiras. Podemos retirar, portanto, do fenômeno particular da auto-ilusão, conseqüências gerais relativas às pretensões narrativistas de oferecerem uma teoria da identidade pessoal alternativa à do modelo neo-lockeano.

Ao entendermos melhor o que exatamente isso quer dizer, ou seja, em que consiste o fracasso da teoria narrativista frente ao modelo lockeano, poderemos entender também uma outra dificuldade interna do narrativismo que é bastante similar à circularidade presente na teoria da memória de Locke. Vejamos em que consiste essa outra dificuldade do narrativismo.

O caráter distintivo da teoria de MacIntyre, se comparada às abordagens lockeana e neo-lockeana, é a descrição do *self*, não, como nas teorias anteriores, apenas em termos mentalistas de relações de memória ou de *quasi*-memória, e sim em termos de relações narrativas, as quais, por força de sua maior expressividade, determinam a necessidade de se levar em conta a dimensão intersubjetiva da constituição do *self*, dimensão essa que inexiste,

ou que, ao menos, não desempenha um papel essencial na teoria lockeana. A dificuldade da teoria de Locke está, como vimos anteriormente, no fato de que ela não pode oferecer um critério não circular da identidade pessoal em termos de relações de memória. Como também já foi visto, em resposta a essas dificuldades, surgem contemporaneamente dois tipos de teorias da identidade pessoal. O primeiro tipo envolve um aperfeiçoamento da perspectiva lockeana, favorecendo uma abordagem impessoal, de terceira pessoa, tal como as teorias causais da memória propostas por Perry e Parfit; o segundo tipo é representado pelas teorias narrativistas, como a de MacIntyre. Visto que as teorias da constituição narrativa do *self* estão comprometidas desde o início com a perspectiva de primeira pessoa e que as teorias do tipo neo-lockeano partem exatamente da premissa oposta, considera-se tradicionalmente que esses dois tipos de teorias são incompatíveis entre si. A visão tradicionalmente aceita é a de que a principal diferença entre as teorias lockeana e neo-lockeana, de um lado, e a teoria narrativista de MacIntyre, de outro lado, reside no fato de que a última adota a perspectiva de primeira pessoa, enquanto que as primeiras adotam a perspectiva de terceira pessoa (por sua vez, como já foi dito, a diferença entre Locke e os neo-lockeanos reside no fato de que, enquanto Locke se baseia estritamente na descrição dos fenômenos que aparecem para um sujeito individual, ao assumir a continuidade psíquica dada pela vivência dos conteúdos da memória como critério de identidade pessoal, os neo-lockeanos, para evitar o caráter circular da teoria de Locke, abandonam essa perspectiva caracteristicamente “fenomenológica” da teoria lockeana, em favor da adoção de uma explicação causal das relações entre a memória e as experiências que elas reportam). Mas a teoria macintyreana da constituição intersubjetiva do *self* também não está isenta de problemas similares aos presentes na teoria de Locke. Particularmente, sua descrição da identidade qualitativa como a condição necessária para o estabelecimento da identidade quantitativa, a partir da noção de unidade narrativa de um personagem, não é satisfatória, como veremos agora.

O problema da circularidade presente na teoria lockeana assume, no contexto da teoria narrativista de MacIntyre, a forma da necessidade de se distinguir personagens reais de personagens meramente ficcionais²⁰. Os personagens ficcionais visados aqui não são meramente os personagens de romances e de peças teatrais, mas também aqueles resultantes de identificações ilusórias que nós mesmos forjamos com personagens que de fato não somos, produzidos por narrativas falsas ou auto-ilusórias. No caso do narrativismo, é preciso caracterizar o personagem real sem pressupor que ele é aquele a que se referem as narrativas reais – ou seja, sem pressupor

²⁰ No caso de Locke, o círculo, como vimos, decorre da indistinção entre memórias reais e memórias aparentes: o único critério de distinção parece ser afirmar que as memórias reais são aquelas referentes a experiências que *eu mesmo* tive no passado – o que, ao invés de definir, pressupõe a identidade pessoal.

que as narrativas reais são aquelas que se referem a experiências reais que *aquele mesmo self* teve no passado (o que faria com que, ao invés de definir, pressupuséssemos a identidade pessoal). É claro que, tanto no caso do lockeano quanto no caso do narrativista, as representações falsas sobre mim mesmo (“eu sou aquele mesmo que lembro ter estado no cinema ontem”; “eu sou aquele mesmo que em minhas narrações de vida apareço como tendo um filho não drogado”) são, em certo sentido, constitutivas do que eu *sou* (eu *sou* aquele que tem tais memórias falsas ou aquele que faz tais narrativas falsas). Mas os *conteúdos* dessas representações não podem ser aquilo que, sem mais, garante minha identidade numérica com a pessoa que elas me apresentam.

O critério utilizado por MacIntyre para a distinção entre personagens reais e ficcionais pode ser depreendido da seguinte passagem de *Depois da virtude*: “A diferença entre personagens imaginários e reais não está na forma narrativa do que eles fazem; está no grau de sua autoria daquela forma e de seus próprios feitos” (MacIntyre [1985] p. 215). MacIntyre tem em mente, nesse trecho, um critério para distinguir personagens reais de personagens fictícios, entendendo esses últimos no sentido usual de personagens de romances ou de peças teatrais, e não no sentido por nós discriminado mais acima, no qual “personagens fictícios” eram entendidos como aqueles constituídos por representações falsas sobre mim mesmo. Apesar de ser nesse segundo sentido que o problema da circularidade, similar ao problema lockeano, é formulado, podemos, ainda assim, partir por ora da distinção entre personagens reais e personagens ficcionais entendida no primeiro sentido (por ex., a distinção entre Shakespeare e Hamlet), pois, se o narrativismo tem razão em utilizar o conceito de personagem real para explicar o que é uma pessoa, ele deve poder distingui-lo do conceito de personagem ficcional em qualquer dos dois sentidos mencionados acima. Vejamos então se a noção de “graus de autoria” dá uma resposta satisfatória para a pergunta sobre a natureza dos personagens reais.

Deixemos de lado por ora alguns dos aspectos obscuros que envolvem a problemática introdução da noção de “grau de autoria”, que sugeriria a existência de entidades ontologicamente escandalosas, já que elas seriam, de acordo com o seu nível de capacidade de auto-produção, um pouco mais semi-reais ou um pouco mais semi-ficcionais. Sem considerar essas dificuldades, tentemos entender o que MacIntyre pode estar querendo assinalar com tal noção.

Em primeiro lugar, poderíamos pensar que a noção de “grau de autoria” visa a tese de que a atividade narrativa do autor, entendido sempre como co-autor, deve atender a determinados limites, impostos pela narrativa maior da comunidade na qual ele está inserido. “Certamente, assim como eles não começam quando querem, eles não podem continuar exatamente como querem; cada personagem é constrangido pelas ações dos outros e

pelos cenários sociais pressupostos em suas ações e nas deles” (MacIntyre [1985] p. 215). Ou seja, um autor seria “mais ou menos” o autor das narrativas que o constituem como um personagem real. Poderíamos criticar a tese da co-autoria pela comunidade, mostrando que seu papel não deve ser exagerado e que o autor individual tem sempre uma prioridade sobre os limites histórico-sociais impostos por suas circunstâncias, mas por ora já é possível criticar essa interpretação da noção de “grau de autoria” simplesmente pelo fato de ela não discriminar, de forma alguma, personagens reais de personagens ficcionais, devido ao fato de que o importante aqui, para a solução do problema, é o esclarecimento de como se dá a relação de constituição entre o autor real e o personagem (real ou ficcional), e não a eventual co-participação nessa constituição. Não deve ser a isso que MacIntyre está fazendo referência neste ponto com tal noção. Uma segunda maneira mais promissora de interpretar a noção de “grau de autoria” neste contexto é dizer que o que distingue o personagem real do personagem ficcional é a presença ou ausência da capacidade reflexiva. Um personagem real seria aquele que tem real capacidade de autodeterminação, ao passo que a autonomia dos personagens ficcionais é ela mesma apenas ficcional (como quando é dito: “Então Ivan Karamazov decidiu ir até a presença do juiz...”). É estranho falar de “graus” nesse sentido, pois o “grau” de autoria dos personagens ficcionais é, propriamente falando, zero. Mas essa parece ser a melhor explicação do que é visado por MacIntyre neste ponto.

Mesmo, porém, que aceitássemos provisoriamente a figura do “personagem real”, ou seja, mesmo que concedêssemos ao narrativista que somos personagens reais de narrativas reais, opondo a isso personagens ficcionais (quer eles pertençam a romances e a peças teatrais, quer pertençam a um modo falso de narrar nossa própria vida), o modelo narrativista se veria enredado na mesma dificuldade da circularidade que parece comprometer a teoria de Locke. Enquanto, para esse último, a única maneira de distinguir memórias reais de memórias aparentes seria apelar para o fato de que minhas memórias reais se referem a experiências que *eu mesmo* tive no passado (o que seria circular), para o narrativista, a única maneira de distinguir narrativas reais de narrativas ficcionais sobre minha própria vida²¹ (como as que ocorrem, por exemplo, nos casos de auto-ilusão) seria apelar para o fato de que minhas narrativas reais se referem a experiências que *eu mesmo* tive no passado (o que também seria circular). Daí se segue, imediatamente, que a única maneira de identificar um personagem real como sendo o mesmo é supor que ele é constituído pelo mesmo autor, ou seja, pela mesma pessoa. Para tanto, é necessário que se identifique a pessoa ao autor, e não mais ao personagem. Não haveria, ao que parece,

²¹ A distinção entre personagens reais e personagens ficcionais, entendendo-se esses últimos como personagens de romances ou do teatro, é trivial, e não precisa ser tematizada explicitamente aqui.

nenhum critério independente de atribuição de identidade – mesmo a continuidade física de um mesmo corpo humano vivo seria compartilhada entre os personagens reais e os ficcionais (entendendo esses últimos, é claro, não no sentido em que Hamlet é um personagem ficcional, mas sim como sendo frutos de narrativas falsas que eventualmente fazemos sobre nossa própria vida). Seria possível talvez reformular o argumento narrativista de modo similar ao que os neo-lockeanos reformularam o argumento original de Locke, introduzindo a idéia de conexões causais de conteúdos de memória, mas veremos abaixo que mesmo essa reformulação neo-lockeana não se livra da acusação de circularidade. *Mutatis mutandis*, seria possível provar também que um argumento similar dos narrativistas é igualmente refutado.

Por ora, basta concluir o seguinte: o narrativismo, ao tentar responder o problema da circularidade presente na teoria lockeana, tem de partir dos termos que são equivalentes, em sua teoria, aos termos do vocabulário psicológico do lockeanismo: ao invés de memórias verdadeiras e memórias falsas, personagens reais e personagens ficcionais. Mas não foi possível (a não ser como uma concessão, expressa no parágrafo anterior²²) nem ao menos começar a examinar como esse problema se recolocaria no contexto do narrativismo, pois a própria distinção inicial entre personagens reais e personagens ficcionais se revelou como não sendo bem fundada (ao contrário do que ocorre na oposição bem fundada entre memórias reais e memórias falsas, que têm conceitos claros). Isso nos leva a constatar que o narrativismo de Macintyre não satisfaz o segundo requerimento de toda teoria da identidade pessoal, já que ele não se mostrou capaz de explicar a auto-ilusão, e, sendo assim, pelo fato de que, ao falhar nesta tarefa, ele se mostrou incapaz de satisfazer o primeiro requerimento, o de prover um critério confiável de distinção entre pessoas e ficções, não podemos considerá-lo como uma teoria correta da pessoalidade.

2.3 - A auto-ilusão e a circularidade do modelo neo-lockeano

A crítica que farei ao modelo neo-lockeano consiste em mostrar que o conceito de auto-ilusão depende internamente daquilo que chamarei de “autoria”, (ou seja, de uma decisão endógena de me auto-iludir) partindo da premissa, aceita pelo neo-lockeano, segundo a qual somos uma conexão de estados mentais conectados causalmente pela memória. Para se determinar essa dependência analítica, é necessário que se faça uma nova elucidação conceitual da noção de auto-ilusão. Para começar, devemos explicar que tipo de relação existe entre a decisão de se auto-iludir e as conexões psicológicas.

²² “Mesmo, porém, que aceitássemos provisoriamente a figura do ‘personagem real’...”.

Uma vez me percebendo como vítima da auto-ilusão, eu posso, ao menos em princípio, retornar, através da memória, àquela ocasião onde decidi – pré-reflexivamente – me auto-enganar. Neste caso, a decisão pré-reflexiva de me auto-iludir parece ser perfeitamente explicável em termos de conexões psíquicas, podendo ser descrita na perspectiva de terceira pessoa como resultado de meras relações entre estágios de pessoas conservando-se a continuidade psíquica. Assim descrita, a decisão de me auto-iludir é apenas mais um dentre os vários tipos possíveis de conexões psíquicas que descrevem minha identidade pessoal. Há, assim, uma relação de constituição entre minha decisão passada e as conexões psicológicas que lhe sucederam, ou seja, parece ser perfeitamente possível explicar minha decisão como consistindo em uma conexão psicológica, ou, de forma mais rigorosa, como tendo sido constituída como uma conexão psicológica.

Até aqui não há surpresas, pois, ao que parece, não haveria nenhum problema em caracterizar a auto-ilusão em termos de terceira pessoa, eliminando a necessidade de introduzir nesta explicação a figura do autor, e, assim, a perspectiva de primeira pessoa. No entanto, o que foi analisado até aqui é a relação entre a decisão de me auto-iludir e as conexões psíquicas na dimensão temporal do passado, ou seja, quando o fenômeno da auto-ilusão, propriamente dito, já não mais existe (estou supondo justamente que é possível voltar-se reflexivamente – isto é, sem ilusões – para aquele momento passado). Mas o que é interessante no fenômeno da auto-ilusão é o fato de que sua definição envolve necessariamente a dimensão prospectiva, pois trata-se de uma decisão de lembrar ou de esquecer, no futuro, conexões psicológicas que constituem agora aquele que presentemente sou, fazendo com que aquilo que se me afigura agora como razões para agir, atue futuramente como causa de minha ação. Considerando a decisão de me auto-iludir nessa perspectiva de primeira pessoa dirigida para o futuro, veremos que o modelo neo-lockeano passa a apresentar problemas incontornáveis²³.

A possibilidade de que uma decisão que tomo agora venha a constituir conexões psicológicas futuras não é algo que possa embaraçar o neo-lockeano, pois seu modelo de identidade pessoal pode muito bem contemplar essa possibilidade, sem que isso produza uma contradição, pois não há nada de contraditório em, por exemplo, uma pessoa programar-se para executar mais adiante uma tarefa e executá-la de fato no futuro. No entanto, o que o neo-lockeano não pode admitir é a possibilidade de que decisões prospectivas produzam conexões psicológicas *falsas*, o que é justamente o caso quando levamos em conta o fenômeno da auto-ilusão. Isso remonta à dificuldade inicial de Locke, formulada por seus críticos diretamente a ele, e que o modelo neo-lockeano busca superar, segundo a qual, como foi anteriormente visto, haveria uma circularidade no critério da memória proposto por Locke

²³ Ou seja, uso como método aqui a recomendação de Korsgaard [1996] de descrever os fenômenos relativos à identidade pessoal a partir da perspectiva do agente, e não da perspectiva de terceira pessoa, de quem examina um “dado” objetivo.

para explicar a identidade pessoal. No entanto, mesmo quando adotamos o critério mais sofisticado de continuidade psicológica de Perry e Parfit, a circularidade reaparece, quando se leva em conta o fenômeno da auto-ilusão. O argumento que se segue pode, portanto, ser considerado como uma reedição da crítica original de Butler a Locke, com as devidas modificações, para abarcar também os neo-lockeanos.

Segundo a definição neo-lockeana de identidade pessoal, essa última consiste na continuidade psicológica produzida pelo tipo correto de causa, ou seja, sem que haja distorções externas. Um estágio temporal da pessoa A pode ser dito *a mesma pessoa* que o estágio temporal de pessoa B se B representa estados ou ações de A e se essa representação de B foi causada por A – e não por fatores externos, como hipnose ou alucinações. Ou seja, a atribuição de identidade pessoal depende de *fatores intencionais* (representações e crenças de B acerca de A) e de *fatores causais* sem distorções externas. No entanto, nada é dito sobre a possibilidade de que existam distorções internas ou autorais. Mas a auto-ilusão é exatamente esse tipo de distorção interna; deste modo, como a auto-ilusão consiste em uma decisão prospectiva de se produzir conexões psicológicas *falsas*, é necessário que haja um critério de identidade do auto-iludido, ou seja, que nos possibilite *identificar como sendo a mesma* a pessoa A, que decide se auto-iludir, e a pessoa B, que é constituída por conexões psicológicas falsas. Mas isso, como veremos, é impossível nesse modelo.

Tomemos um exemplo: digamos que eu perceba estar com uma doença grave que me levará à morte em poucos meses; o medo da morte e o desejo de viver, porém, causam minha progressiva conversão à crença oposta: estou bem, apenas um pouco cansado, recuperando-me de uma doença que já começa a passar. Descrevendo isso em terceira pessoa, temos que o estágio temporal A é dotado de certas crenças e desejos (entre elas a crença de que está gravemente doente, que simbolizaremos por *p*) que causam, em um momento posterior, e sem que tais causas apareçam mais em seu aspecto intencional para o próprio sujeito que as tem, a crença, presente no estágio temporal B, de que $\sim p$. A princípio, não há dificuldades em descrever isso segundo o aparato conceitual neo-lockeano: uma crença falsa, que pode ser identificada enquanto tal em terceira pessoa (é possível saber que tal indivíduo está se auto-iludindo, e que de fato sua crença de que $\sim p$ não corresponde aos fatos, a saber, neste caso, ao fato de que morrerei em poucas semanas), foi causada por um desejo que não se apresenta mais conscientemente enquanto tal, mas que, também em terceira pessoa, pode ser descrito como a causa oculta daquela crença. Entretanto, um exame mais cuidadoso mostra o fracasso desse modelo. O que B representa sobre si mesmo é que ele é aquela pessoa que nunca esteve gravemente doente, que está agora em recuperação e que terá muitos anos de vida pela frente. Ele é essa pessoa? Não, isso é uma ilusão, uma auto-ilusão. Seria como se eu acreditasse que sou aquela mesma pessoa que esteve ontem no cinema, sem saber que fui hipnotizado para acreditar falsamente que ontem estive

no cinema. Mas há uma diferença entre esses dois casos: enquanto no caso do hipnotizador uma causa externa interveio e me fez ter uma representação falsa acerca de mim mesmo, eliminando assim a atribuição correta de identidade, no caso da auto-ilusão não há nenhuma causa externa que tenha causado a falsa representação do que sou. Ou seja, segundo os próprios critérios neo-lockeanos, eu teria de dizer que *sou* aquela pessoa que acredita que nunca esteve doente; afinal, eventos mentais conectados endogenamente causaram a representação em questão. Mas como é possível então falsificar essa representação, dizer que não sou aquilo que acredito ser? O único critério fornecido pelos neo-lockeanos, como vimos, é a existência de uma rede de conexões causais na mente, o que foi plenamente cumprido nesse caso. Mas isso gera um absurdo, porque *sabemos* (em terceira pessoa, descrevendo “de fora” tal situação) que essa crença é irracional e não corresponde aos fatos empíricos: eu de fato morrerei em algumas semanas, apesar de querer desesperadamente que isso seja falso. Em outras palavras, eu *não sou* aquela pessoa que acredita que nunca esteve doente; antes, sou aquela que se auto-ilude para escapar ao enfrentamento direto da morte.

Mas poder-se-ia objetar: eu não posso ter crenças falsas? Corriqueiramente, é verdade dizer que “causo” em mim mesmo falsas opiniões, por exemplo: quando, por desatenção, passo a acreditar que o nome de minha vizinha é Maria, quando de fato é Maia. Mas o caso da auto-ilusão é diferente: não se trata de um erro sobre fatos externos do mundo, mas de uma representação inadequada sobre mim mesmo, o que, pela teoria neo-lockeana, não poderia ser explicado, uma vez que houve um nexos causal endógeno ligando o estágio A ao estágio B. Poder-se-ia de novo objetar: mas de fato *não sou* aquela pessoa tal como é representada por B; trata-se de uma crença empiricamente falsa sobre mim mesmo. Entretanto, o que eu sou, segundo o modelo neo-lockeano, *além* de uma série de representações e desejos conectados por causas endógenas? Nada. Se digo que de fato não sou aquilo que é representado por B, então teria de dizer que não sou redutível *àquela* série da qual B faz parte; sou *outra* coisa (outra série?). Mas isso significa justamente que o modelo neo-lockeano é falso, pois ele afirma que sou redutível *àquela* série.

Resumindo: a objeção de Butler a Locke foi, assim, retomada para o próprio neo-lockeano, de forma que podemos descartar também essa versão revisada de reducionismo mentalista da identidade pessoal. A solução do neo-lockeano para o problema teria de mostrar que as conexões psicológicas que constituem B são aquelas que foram pré-programadas por A, mas isso é já supor que A e B são a mesma pessoa, o que é supor aquilo que se quer provar, gerando-se um círculo lógico. Como a identidade pessoal é definida em termos de continuidade psicológica, então não há, nesse modelo, um critério não-circular de identidade pessoal; assim sendo, é possível agora dizer que o modelo neo-lockeano é falso.

Em outras palavras: o modelo neo-lockeano não cumpre o segundo requerimento de toda teoria da pessoalidade; logo, ele também não é capaz de

cumprir o primeiro requerimento. Assim sendo, este modelo não pode ser considerado uma explicação correta do conceito de pessoa.

3. Conclusão:

Reconstrução do narrativismo ou ceticismo tanto com respeito à possibilidade de caracterização narrativista da noção de pessoa quanto com respeito à possibilidade de caracterização da noção de agente moral a partir da ética das virtudes

Levando-se em conta toda a discussão anterior, com respeito às dificuldades internas tanto do narrativismo de MacIntyre quanto do modelo neo-lockeano proposto por Parfit, o desenvolvimento da pesquisa deverá se pautar em duas alternativas possíveis e opostas em relação aos conceitos de pessoa e de agente moral; a primeira delas seria a reconstrução da teoria narrativista da pessoalidade, para além de MacIntyre, assumindo-se o ônus de produzir uma definição não-circular de identidade pessoal que a torne compatível com a possibilidade do auto-engano, possibilitando assim a caracterização do agente moral a partir de uma ética das virtudes e retomando o projeto macintyreano; porém, no caso do fracasso dessa tentativa, restaria a segunda alternativa, que seria a adoção de uma postura cética com respeito à possibilidade de uma caracterização narrativista do conceito de pessoa, bem como com respeito à possibilidade da caracterização do agente moral a partir da ética das virtudes. Tal postura cética se justificaria, no caso do fracasso da tentativa de reconstrução do narrativismo, em razão de que, por um lado, a plausibilidade inicial do narrativismo (da qual tratarei adiante), compatível com muitas visões do senso comum pré-filosófico contemporâneo sobre a natureza das pessoas, nos revela que o conceito de pessoa está intimamente vinculado com o de narrativa, e porque, por outro lado, as tentativas narrativistas para a caracterização desse conceito fracassaram. Ou seja, se o narrativismo é nossa melhor explicação para o conceito de pessoa e, posteriormente, para o conceito de agente moral, e se essa doutrina nos enreda em paradoxos, talvez a melhor atitude seja assumir de vez uma postura cética quanto à possibilidade de uma caracterização narrativa dos conceitos de pessoa bem como quanto à possibilidade de uma conceituação do agente moral a partir da ética das virtudes.

Apresento aqui, para finalizar, uma razão que podemos usar como justificativa para crença na necessidade da reconstrução do narrativismo (ou, ao contrário, para a atitude cética que redundaria do fracasso de uma tal reconstrução), qual seja, a plausibilidade inicial do narrativismo. Se julgarmos a partir do senso comum, podemos dizer que as teses do narrativismo

são possuidoras de uma plausibilidade inicial²⁴. A plausibilidade inicial do narrativismo é de tal ordem que podemos concluir desde já que algo nessa teoria deve ser aceito como verdadeiro por qualquer teoria sobre a identidade pessoal, mesmo que o narrativismo deva ser modificado por considerações adicionais e mesmo que não mantenhamos o senso comum como critério de correção de nossas teses metafísicas sobre a pessoalidade. O que verificamos cotidianamente, quando observamos o desenvolvimento de crianças até a fase adulta é, entre outras coisas, uma verdadeira batalha infantil por reconhecimento e cuidado dos adultos, e também, por outro lado e de forma conflituosa, por independência desses mesmos adultos. Desenha-se, a partir desse conflito de necessidades, uma pequena trama, que, de fato, representa o núcleo dramático de toda e qualquer vida. Essa trama manifesta-se em uma necessidade cada vez maior de a criança narrar sua própria vida, necessidade essa que se impõe à medida que as capacidades cognitivas ligadas à linguagem se desenvolvem paulatinamente. Dessa forma, parece razoável dizer que a vida humana, mesmo se vista por olhos desarmados de teorias religiosas, psicológicas, sociológicas ou filosóficas, se apresenta, como que “espontaneamente”, em termos de um jogo dramático, ou de uma peça teatral. Esta é, portanto, uma razão, ainda que pré-teórica, para acreditar-se na necessidade da reconstrução do narrativismo ou, ao invés, para adotarmos uma postura cética tanto quanto à possibilidade de uma caracterização narrativa dos conceitos de pessoa, como quanto à possibilidade da caracterização dos agentes morais a partir da ética das virtudes.

Bibliografia

- AUDI, Robert. Self-deception, Action, and Will. In *Erkenntnis* 18 (1982): 133-158.
- BUTLER, Joseph. Of Personal Identity. Primeiro Apêndice de *The Analogy of Religion* (1736). In PERRY, John (org.). *Personal Identity*. Berkeley / Los Angeles / London: University of California Press, 1975, p. 99-106.

²⁴ Em que consiste essa plausibilidade do narrativismo? Ela fica mais clara se propusermos uma espécie de “experimento imaginário”, a saber: o que representaríamos se tentássemos nos representar uma pessoa adulta e dotada de todas as suas faculdades mentais e físicas, mas que fosse incapaz de narrar sua própria vida? Alguém poderia dizer: eu posso me representar um eremita adepto do voto de silêncio, que vive isolado da sua comunidade de origem, contemplando a natureza ou a divindade. Mas esse claramente não seria um bom exemplo de ausência de *capacidade* de narrar sua própria vida; ao contrário, é preciso se engajar em uma narrativa bastante complexa (inicialmente em comunidade, depois talvez em um solilóquio) para se adotar tal estilo de vida. A necessidade de contar o que nos aconteceu e de antecipar narrativamente o que nos acontecerá, para nós mesmos e (na maioria dos casos) para os outros, parece fazer parte de tal modo do que *somos* que é extremamente plausível supor que há um vínculo conceitual entre a noção de pessoa e sua capacidade narrativa, ou seja, que não é acidental que narremos nossa própria vida. Trata-se de uma capacidade que deve ser exercida se é que somos (ou nos tornamos) pessoas.

DAVIDSON, Donald. Paradoxes of Irrationality. In WOLHEIM, A. / HOPKINS J. (ed.). *Philosophical Essays on Freud*. Cambridge: Cambridge University Press, 1982.

DENNETT, Daniel. *Consciousness Explained*. Boston: Little / Brown, 1991.

_____. The Self as a Center of Narrative Gravity. In: KESSEL, F. / COLE, P. / JOHNSON, D. (org.). *Self and Consciousness: Multiple Perspectives*. Hillsdale (NJ): Erlbaum, 1992.

FONSECA, José Sérgio D. da. Manipulação genética e a crise da identidade moderna: Taylor, Dennett e o 'naturalismo tardio'. In *Síntese* 31 (2004): 65-90.

KORSGAARD, Christine. Personal Identity and the unity of agency: A Kantian response to Parfit. In ID. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996, p. 363-397.

LOCKE, John. *An Essay Concerning Human Understanding*. Edição e introdução de Peter H. Nidditch. Oxford: Oxford University Press, 1979.

MACINTYRE, Alasdair. *After Virtue. A Study in Moral Theory*. 2ª edição revista. London: Duckworth, 1985 (1981).

MELE, Alfred R. Real Self-Deception. In *Behavioral and Brain Sciences* 20 (1997): 91-136.

PARFIT, Derek. *Reasons and Persons*. Oxford: Clarendon Press, 1984.

PERRY, John (org.). *Personal Identity*. Berkeley / Los Angeles / London: University of California Press, 1975.

PINHEIRO, Ulysses. *Fraqueza da vontade e auto-ilusão segundo Espinosa*. Manuscrito não publicado.

SARTRE, Jean-Paul. *L'être et le néant. Essai d'ontologie phénoménologique*. Paris: Éditions Gallimard, 1943.

_____. *O ser e o nada. Ensaio de ontologia fenomenológica*. Tradução e notas de Paulo Perdigão. Petrópolis: Vozes, 1997.

Endereço do Autor:
Rua Nascimento Silva, 110 – apto. 201
22421-030 Rio de Janeiro – RJ
e-mail: sergio_fonseca@uol.com.br