O Papel das Evidências Empíricas na Filosofia Política Contemporânea

E algumas de suas implicações

Universidade Federal do Rio Grande do Sul
Instituto de Filosofia e Ciências Humanas
Programa de Pós-Graduação em Filosofia

O Papel das Evidências Empíricas na Filosofia Política Contemporânea

E algumas de suas implicações

Tese apresentada ao Programa de Pós Graduação em Filosofia da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do grau de Doutor em Filosofia

Daniela Goya Tocchetto

Orientador:
Dr. Nelson Boeira

Porto Alegre - RS

Março de 2014

Dedicatória

Para minha irmã.

Agradecimentos

Deixo aqui meu sincero ‘muito, muito obrigada’ a todos aqueles que de alguma maneira contribuíram para este trabalho.

Ao meu orientador, Nelson Boeira, que me ajudou em todas as etapas desse caminho e que, da maneira mais bonita e humana, esteve do meu lado e ‘segurou todas as pontas’ quando eu mais precisei.

A todos os meus professores, pelos preciosos ensinamentos.

Aos colegas e amigos do grupo de pesquisa em Filosofia, Economia e Direito, pelos encontros e discussões que em muito enriqueceram as minhas pesquisas.

Ao professor André Klaudat, em especial, pelo admirável exemplo de postura acadêmica e pelas inúmeras horas de auxílio nas leituras de David Hume.

À Capes, pelo apoio financeiro sem o qual esse trabalho não teria se tornado concreto.

Aos meus amigos e à minha família, pelo amor e pela paciência.

Ao meu marido, Thomas, por tudo. Sempre.

> Rawls does not conceive of moral philosophy as depending primarily on the analysis of valid moral argument. Rather, he thinks of a theory of justice as analogous to a theory in empirical science. It has to square with what he calls ‘facts’, just like, for example, physiological theories. *But what are the facts?*
>
> Hare, 1973, p. 145

Resumo

A presente tese é composta por quatro artigos que, embora relativamente independentes, foram escritos tendo em vista um objetivo comum. Este objetivo comum, fio condutor do trabalho, é a defesa da ampliação do uso de evidências empíricas concernentes ao nosso comportamento moral no desenvolvimento de teorias contemporâneas de justiça. Além dessa defesa, o trabalho discute duas implicações relevantes de um uso adequado dessas evidências pelos filósofos políticos. De antemão, é importante esclarecer que este objetivo não equivale à afirmação de que os filósofos políticos contemporâneos são completamente indiferentes aos resultados das ciências empíricas. De maneira análoga, também não equivale à completa desconsideração de sua metodologia atual. Feitas essas ressalvas, eu me concentro nas seguintes questões nos quatro artigos que compõem esta tese. No primeiro artigo, eu apresento os principais argumentos contrários a uma incorporação mais profunda de evidências empíricas na filosofia política contemporânea e, em seguida, exponho e discuto um rol de razões suficientes para a desconsideração desses argumentos. No segundo artigo, após ter estabelecido a maneira própria de colaboração entre as ciências empíricas e a filosofia politica, eu apresento uma extensa revisão da literatura empírica existente sobre intuições, crenças e comportamentos relacionados com os conceitos de justiça e equidade. Esta revisão inclui as pesquisas mais significativas sobre o nosso comportamento moral realizadas nas últimas três décadas nas áreas de primatologia, biologia evolutiva, economia experimental, psicologia moral, psicologia política e social, e neurociência. Por fim, nos dois últimos artigos, eu discuto duas implicações importantes de uma filosofia política empiricamente informada. No terceiro artigo, eu busco recuperar o sentimentalismo moral na filosofia política, argumentando que a primeira lição que devemos extrair das evidências empíricas discutidas no artigo anterior é que a moralidade é tanto uma questão de sentimentos quanto de razões. Finalmente, no quarto artigo, eu defendo que uma segunda implicação importante de uma filosofia política empiricamente informada é o ressurgimento de princípios de merecimento em teorias de justiça distributiva. De forma a colaborar com esse ressurgimento, eu realizo nesse último artigo um experimento que investiga as intuições da população em geral sobre diferentes bases de merecimento. De tal modo, eu espero contribuir para um melhor entendimento das nuances desse importante conceito.

Abstract

The present dissertation consists of four nearly self-contained articles written with a common goal, namely, the investigation of the proper role of empirical evidence in contemporary political philosophy and of some of its implications. At the outset, it is important to clarify that this common goal does not amount to stating that contemporary political philosophers have been completely indifferent to the results of the empirical sciences. Neither does it amount to a plea for dismissing their current methodology, replacing it for some entirely new way of conducing the development of theories of justice. In this vein, I focus on the following issues in the four papers that compose this dissertation. In the first paper I address the main arguments that have been presented against a deeper incorporation of empirical evidence in contemporary political philosophy, along with the reasons for the dismissal of these arguments. In the second paper, after the grounds have been settled for a proper collaboration between the empirical sciences and normative political philosophy, I present an extensive review of the current empirical literature on human intuitions, beliefs, and behaviors related to the concepts of justice and fairness. This review includes the most significant research involving these concepts during the past three decades in the areas of primatology, evolutionary biology, experimental economics, moral psychology, political and social psychology, and neuroscience. My hope is that making all these novel research programs and some of its interesting findings easily available for political philosophers will fuel the development of an empirically informed practice. At last, in the two final papers, I discuss two important implications of an empirically informed political philosophy. In the third paper, I undertake the ambitious task of reclaiming moral sentimentalism in political philosophy. I claim that acknowledging that human morality is as much a matter of sentiments as it is a matter of reason is the first important lesson we can learn from the empirical evidence portrayed in the preceding paper. Finally, in the fourth paper, I claim that a second notable implication of taking empirical evidence seriously is the resurgence of principles of desert in theories of distributive justice. In an attempt to build on this resurgence, I propose and implement an experiment that investigates the folk's intuitions on different basis of desert.

Apresentação

A presente tese é composta (como faculta a Resolução número 093/2007, da Câmara Pós Graduação da UFRGS) de quatro capítulos redigidos em inglês, os quais serão posteriormente submetidos à publicação como artigos separados. O título, esta Apresentação, e o Epílogo estão redigidos em português, respeitando as exigências para uma tese nesse formato. Apesar de relativamente autossuficientes, os capítulos que compõem essa tese foram redigidos tendo em vista um objetivo comum: a defesa de uma maior utilização de evidências empíricas na filosofia política contemporânea, bem como a discussão de algumas das implicações advindas de tal uso. A tese central é que filósofos políticos contemporâneos devem considerar de forma mais séria e comprometida, quando do desenvolvimento de suas teorias, os resultados empíricos das ciências naturais e sociais sobre o comportamento moral humano. A partir dessa tese, segue que tal consideração tem como consequência o abandono de uma perspectiva *estritamente* racionalista, no estilo Kantiano, em prol de uma perspectiva mais próxima do sentimentalismo moral, como inicialmente desenvolvido por David Hume e Adam Smith.

***

A filosofia política constitui um campo de investigação filosófica preocupado, sob uma ótica ampla, com o estudo das organizações sociais humanas. De maneira mais específica, o objetivo do filósofo político é a elaboração de um conjunto de princípios capazes de guiar o modo como vivemos não sob uma perspectiva individual atomística, mas sim como membros ativos de uma comunidade cooperativa. As questões com as quais um filósofo político se defronta dizem respeito à maneira através da qual devemos compreender nossas responsabilidades recíprocas enquanto cidadãos; a qual o tipo de tratamento que um ser humano deve ao outro enquanto cidadão em uma sociedade.

Na busca pelos princípios capazes de guiar corretamente as nossas instituições sociais, a principal virtude na qual os filósofos estão interessados é a *Justiça* – definida por Rawls (1971) como a virtude primeira das instituições. E, na busca pela justiça, uma das principais preocupações da filosofia política são as questões levantadas pela chamada *Justiça Distributiva*.

A principal função de princípios de justiça distributiva é guiar a condução das instituições sociais no que diz respeito à alocação das vantagens e desvantagens decorrentes da vida em sociedade, tais como impostos, tratamento médico, educação, etc. Uma teoria de justiça distributiva é fundamental para o desenvolvimento de uma sociedade – ainda que seus princípios estejam presentes de maneira apenas tácita entre seus membros, como no caso de sociedades em pequena escala. Do mesmo modo, para que princípios de justiça sejam bem-sucedidos, é necessário que eles sejam capazes de "persuadir todas as pessoas a regular seu senso de justiça intuitivo de acordo com esses princípios" (Miller, 2003, p. 21; tradução própria).

Não obstante o papel central ocupado por teorias de justiça distributiva no debate político-filosófico contemporâneo, ainda não se atingiu nada próximo de um consenso acerca de quais princípios devem ser adotados na alocação de recursos sociais e econômicos entre os membros de uma sociedade. Nesse sentido, David Miller (2003) ressalta um aspecto ainda mais preocupante do atual cenário,

> (...) a filosofia política e moral contemporânea, de cunho liberal, nos apresenta um espetáculo de desacordo profundo e contínuo entre teorias de justiça alternativas. Cada teoria sustenta ter revelado de maneira irrefutável a verdade, mas não há razão para acreditar que essa competição entre diferentes teorias será um dia resolvida. (p. 112; tradução própria)

Em face da magnitude da relevância do objeto de análise da filosofia política e do presente estado de completo desacordo sobre qual teoria de justiça é apropriada para reger nossas instituições, cabe a pergunta: quais as razões que explicam termos atingido tal estado de discordância? Não restam dúvidas de que uma grande parte da resposta para essa pergunta pode ser encontrada na complexidade inerente ao tema da justiça. Entretanto, atribuir o atual estado de dissenso acerca de princípios de justiça alocativa inteiramente a essa complexidade seria não apenas um equívoco, mas a própria admissão da impossibilidade de obtenção de acordo (mínimo!) sobre tal matéria.

Uma alternativa mais promissora, a meu ver, é iniciar pela investigação de um notável aspecto comum à filosofia política contemporânea, a saber, o seu método de estudo. Esse método é conhecido de acordo com terminologia não-formal como *armchair philosophy*, e se refere ao método de abstração a partir de intuições pessoais do filósofo. A ideia básica reside em começar uma teoria a partir de hipóteses elaboradas individualmente pelo filósofo sobre quais são as intuições básicas das pessoas sobre determinado tema e, a partir dessas hipóteses não-empiricamente testadas, construir um edifício teórico através da argumentação racional de forma a derivar um conjunto de princípios abstratos. Os filósofos empiristas, tendo em Hume seu expoente, já apontaram diversas falhas inerentes a esse tipo de metodologia. No entanto, a grande maioria dos filósofos políticos contemporâneos segue a tradição racionalista kantiana.

Esse caminho metodológico tornou os filósofos políticos contemporâneos menos propensos à utilização dos resultados recentes que vêm emergindo das ciências empíricas sobre o nosso comportamento moral – principalmente os resultados que vem sendo revelados pelas ciências naturais. Esse descaso com o que é empírico poderia ser considerado apropriado caso existissem razões consistentes para a impossibilidade de colaboração entre aquilo que é empírico e aquilo que é normativo. Não obstante, tais razões inexistem enquanto consistentes. Pelo contrário, o que encontramos é um rol de razões[1] que sugerem a adoção de uma metodologia não alheia a resultados empíricos; uma metodologia capaz de incorporar de maneira adequada os avanços das ciências tanto sociais quanto naturais no âmbito da moralidade humana.

A dificuldade de inclusão do universo do empírico pela filosofia política contemporânea carrega consequências que ultrapassam as acusações de inexistência de razões consistentes. Como eu vou argumentar, essa negação traz como consequência mais grave a incapacidade do reconhecimento da natureza sentimentalista da nossa moralidade pelos filósofos políticos contemporâneos. Essa é uma consequência séria na medida em que essa incapacidade de compreender de maneira acurada a natureza de nossos julgamentos morais pode conduzir à desconexão entre teoria e realidade, com altos custos em termos políticos, econômicos e sociais.

---

[1] Essas razões constituem o foco do primeiro artigo que compõe essa tese.

As Razões para a atribuição de um papel mais significativo às ciências empíricas na Filosofia Política Contemporânea

O meu objetivo, qual seja, a defesa da relevância dos resultados empíricos para a filosofia política contemporânea, conduz obrigatoriamente a um exame dos argumentos utilizados pelos filósofos políticos contemporâneos na justificação da sua posição metodológica atual. Essa posição, como mencionado, se caracteriza pelo racionalismo Kantiano e pelo papel secundário atribuído às ciências empíricas no desenvolvimento de teorias normativas. Nesse contexto, o exame desses argumentos constitui o foco do primeiro artigo desta tese.

De acordo com David Miller, os filósofos políticos contemporâneos apelaram sobretudo a dois argumentos a fim de abster-se de *sujar as suas mãos com dados empíricos* (2003, p. 42). O primeiro argumento afirma que a pesquisa empírica é incapaz de revelar os juízos ponderados das pessoas sobre a justiça, ao passo que o segundo argumento baseia-se na diferença lógica entre afirmar como as coisas *devem* ser e afirmar como elas *de fato são*. Esse segundo argumento é amplamente conhecido como *falácia natural*: a impossibilidade lógica de derivar uma assertiva de 'dever' a partir de uma assertiva de 'ser'. Dessa forma, Miller (2003) afirma que a relutância dos filósofos políticos em atribuir às evidências empíricas um papel mais significativo no desenvolvimento de teorias de justiça deriva principalmente de uma distinção entre *justificação* e *aceitação*: mostrar que uma crença é aceita, afirmam os filósofos, não equivale a mostrar que ela é justificada, nem obrigatória.

No primeiro caso, a crítica dos juízos ponderados diz respeito à falta de conhecimento especializado da população sobre a moralidade. Não se trata de uma afirmação sobre a irrelevância da intuição popular para a teorização normativa. Pelo contrário, existe uma longa tradição de dependência das intuições humanas na filosofia moral e política. Por exemplo, filósofos tão distintos quanto Aristóteles e Rawls explicitamente apelaram para intuições de justiça por eles supostas como amplamente aceitas pela população no desenvolvimento de suas respectivas teorias. Assim, o que atualmente impede o filósofo de se utilizar dos resultados empíricos de maneira mais significativa é uma postura metodológica: o entendimento de que a maneira apropriada de proceder para desvelar as intuições morais humanas é o processo de introspecção filosófica. Nesse sentido, a introspecção filosófica seria o

método mais apropriado em face da incapacidade da população em geral de formular corretamente os seus juízos ponderados sobre a moralidade – numerosos filósofos contemporâneos adotam essa mesma atitude metodológica.

No segundo caso, a crítica da falácia natural diz respeito às condições lógicas limitando ou permitindo a colaboração entre as teorias ético-normativas e as ciências empíricas. Nesse sentido, essa segunda crítica refere-se estritamente a uma reivindicação lógica. A Lei de Hume, tal como indicada no *Tratado da Natureza Humana*, afirma que, enquanto o valor lógico de ser verdadeiro ou falso pode ser anexado a assertivas empíricas, o mesmo não é possível para assertivas de natureza normativa. Assim, é logicamente inadmissível inferir afirmações de *dever* a partir de afirmações de *ser*. A questão em jogo aqui, eu irei argumentar, é que não é necessário (nem correto!) negar essa impossibilidade lógica para abraçar uma filosofia política empiricamente informada.

É importante salientar que a defesa de uma compreensão empírica mais ampla do conceito principal da filosofia política, ou seja, da *Justiça*, não implica um reconhecimento ingênuo de crenças aceitas como crenças justificadas. Assim como também não implica uma infração às regras lógicas. O reconhecimento da relevância dos dados empíricos constitui sim o reconhecimento devido de seu papel na construção de teorias políticas que sejam confiáveis e viáveis, como é argumentado no primeiro artigo que compõe esta tese.

Dessa forma, os argumentos a favor de uma consideração séria e comprometida das evidências empíricas para a teorização sobre a justiça constituem o foco desse primeiro artigo. Os argumentos são organizados, de forma a adereçar as duas principais críticas apresentadas pelos filósofos contemporâneos, em dois grupos principais: (i) contra a crítica dos juízos ponderados, e (ii) contra a crítica da falácia natural. Cabe ressaltar, mais uma vez, que esse segundo grupo de argumentos não implica uma refutação da falácia natural; a reivindicação lógica permanece válida. Dessa forma, os argumentos que são expostos nessa segunda subseção são destinados apenas à refutação do uso da falácia natural como um impedimento para a colaboração empírico-normativa.

Após a exposição e análise de todos os argumentos pertinentes, ficará claro que temos um longo rol de razões a favor da incorporação de evidências empíricas no processo de teorização político-filosófica. Citando apenas uma dessas razões, considere a falta de orientação prática fornecida por princípios distributivos

dissociados dos padrões reais do comportamento humano. Existem várias outras razões consistentes para que os filósofos políticos contemporâneos passem a olhar com mais atenção para as evidências empíricas sobre o nosso comportamento moral e, como acima mencionado, elas recebem seu devido tratamento no primeiro artigo que compõe essa tese.

O Caráter Racional das Teorias de Justiça Contemporâneas

A consequência mais marcante da escolha metodológica dos filósofos políticos contemporâneos é a prevalência do racionalismo kantiano como fundamento último das principais teorias de justiça atuais. Com relação às teorias de justiça distributiva, como anteriormente mencionado, existe hoje uma gama de teorias contemporâneas alternativas. Essas teorias apresentam variações em diversas dimensões, tais como: qual bem deve ser o foco da distribuição – renda, riqueza, oportunidades, trabalho, bem-estar, etc.; e (ii) qual deve ser a regra distributiva – igualdade, maximização, livre mercado, etc. Apesar das diferenças, o que é importante ressaltar aqui é que a grande maioria dessas teorias compartilha do racionalismo kantiano. Elas foram edificadas a partir de tijolos racionalistas, e a força normativa de seus princípios deriva, como em Kant, do uso da nossa capacidade racional.

As principais teorias contemporâneas de justiça podem ser divididas em duas categorias amplas: (i) liberalismo igualitário, e (ii) libertarianismo. Por um lado, a preocupação dos libertários repousa exclusivamente sobre a proteção de direitos individuais, tais como vislumbrados inicialmente por John Locke: direitos naturais à vida, à liberdade e à propriedade. Por outro lado, os liberais igualitários defendem uma visão mais inclusiva, acrescentando à importância do respeito aos direitos fundamentais uma preocupação constante com as injustiças geradas pela nossa sociedade.

A Distância dos Resultados Empíricos

As teorias de justiça acima referidas guardam outro notável aspecto em comum – este último, consequência da opção pelo racionalismo. Esse aspecto diz respeito à considerável distância existente entre a natureza da moral como descrita por essas teorias e os resultados que vêm sendo revelados pelas ciências empíricas sobre a natureza da moralidade humana. Dessa forma, a incorporação de evidências empíricas pelos filósofos será capaz de exercer um impacto significativo na nossa compreensão do conceito de justiça. Primatologistas, biólogos evolucionistas, psicólogos morais e sociais, e neurocientistas – dentre outros – vêm revelando nas últimas décadas dados surpreendentes sobre nosso comportamento moral. Esses dados apontam na direção de uma moralidade muito mais ligada a emoções do que poderia ser esperado pelos filósofos neokantianos.

Por exemplo, a literatura empírica relata extensa evidência de que nossos juízos morais são provocados por reações emocionais, e que somos facilmente *enganados* pelas nossas próprias intuições morais (Haidt et al., 1993). Inúmeros outros experimentos mostram que nossos juízos morais são fortemente afetados por estímulos ambientais, heurísticas e vieses, intuições emocionais e outras influências semelhantes (e.g. Cushman et al., 2006; Greene et al., 2004; Sinnott-Armstrong et al., 2010; Wheatley & Haidt, 2005). Isso para mencionar apenas alguns dos resultados de apenas uma das correntes de investigação empírica que não devem mais ser ignoradas pelos filósofos políticos.

Neste contexto, o objetivo do segundo artigo que compõe esta tese é fornecer uma extensa revisão da literatura empírica existente sobre intuições, crenças e comportamentos humanos relacionados com o conceito de justiça. Essa revisão inclui algumas das pesquisas mais significativas envolvendo este conceito, durante as últimas três décadas, nas áreas de primatologia, biologia evolutiva, economia comportamental, psicologia moral, psicologia política e social, e neurociência. O objetivo deste primeiro artigo é duplo: tornar todos estes novos programas de investigação e alguns de seus resultados mais interessantes facilmente disponíveis para os filósofos políticos e, ao fazê-lo, fomentar o desenvolvimento de uma abordagem metodológica interdisciplinar em filosofia política, uma área que por natureza é multidisciplinar – e que deve, portanto, ser tratada como tal.

As Implicações de uma Filosofia Política Empiricamente Informada

Depois de avaliar os argumentos favoráveis e contrários à incorporação de evidências empíricas na teorização sobre justiça e de explorar os novos resultados das pesquisas empíricas sobre o nosso comportamento moral, eu passo então para uma discussão preliminar das possíveis implicações desse debate. Em primeiro lugar, eu defendo que uma consideração séria e comprometida das evidências empíricas irá fomentar uma concepção mais sentimentalista do conceito de justiça. E, em segundo lugar, eu defendo que essa mudança de atitude metodológica terá também como consequência a alteração do status atualmente concedido aos princípios de *merecimento* em teorias contemporâneas de justiça.

(i) Sentimentalismo

Como discutido previamente, as teorias de justiça distributiva vêm no último século relegando estados afetivos a um papel secundário no processo de derivação de seus princípios. Dentre os dois iluminismos que ocorreram no século XVIII, os filósofos políticos contemporâneos – tendo Rawls como seu principal mentor – seguiram de maneira quase exclusiva apenas o iluminismo racionalista (Frazer, 2010). Um exemplo paradigmático do rebaixamento do papel das emoções na filosofia política é o fato de que o próprio Rawls analisa a nossa estrutura afetiva apenas após ter finalizado a construção de uma base racional supostamente sólida para ambos os seus princípios de justiça. Na interpretação de Rawls, nossas emoções desempenham um papel subsidiário nas teorias de justiça. Nesse sentido, Rawls argumenta que o entendimento da nossa estrutura afetiva é relevante apenas enquanto necessário para a manutenção da estabilidade dos princípios previamente estabelecidos. No entanto, como vimos, as ciências empíricas têm nas últimas décadas demonstrado que as nossas regras morais são menos kantianas do que os racionalistas poderiam ter previsto.

É, para dizer o mínimo, surpreendente que, apesar de todas as evidências apontando para uma natureza mais emocional da nossa moralidade, os filósofos políticos contemporâneos sigam alheios ao sentimentalismo moral. Nesse contexto, o

terceiro artigo que constitui esta tese tem como foco, seguindo Frazer (2010), um exame dos principais argumentos utilizados pelos filósofos na rejeição do sentimentalismo moral. Dentre esses argumentos encontramos, por exemplo: (i) o receio de cair em um relato meramente descritivo da moralidade, sem poder normativo, e (ii) o problema da separação das pessoas, definido por Rawls como a afirmação de que a nossa experiência de empatia nos conduz a ignorar a inviolabilidade dos indivíduos. Se a empatia de fato turva a distinção entre indivíduos, o sentimentalismo moral pode realmente ser incompatível com uma teoria liberal da justiça construída em torno do valor da autonomia e dos direitos individuais (Frazer , p. 94, 2010). Todavia, nem o primeiro nem o segundo argumentos apresentam um perigo real para sentimentalistas morais, como será devidamente discutido no terceiro artigo – juntamente com a refutação de duas importantes críticas adicionais ao sentimentalismo moral.

Em seu livro mais recente, *The Enlightenment of Sympathy* (sem tradução para o português), Michael Frazer dá início ao trabalho duro de construção de uma visão mais sentimentalista da justiça. Na mesma linha, o terceiro artigo da presente tese constitui também um esforço nessa direção. No final desse artigo, portanto, irei discutir (de maneira altamente preliminar) algumas das implicações da utilização do sentimentalismo na filosofia política. Só para citar uma implicação notável, recordemos que, atualmente, a deliberação política é centrada na razão e na erradicação das emoções da arena dos debates públicos. No entanto, esta atitude de exclusão das emoções pode ser altamente problemática se de fato nossa natureza moral é descrita de maneira mais acurada por Hume e Smith do que por Kant. A adoção do sentimentalismo moral aponta na direção de uma maior participação da retórica na esfera política, de tal forma a proporcionar o envolvimento devido das emoções apropriadas e, assim, gerar uma melhor condução da vida pública.

(ii) Merecimento

O conceito de merecimento praticamente desapareceu da filosofia política contemporânea desde Rawls e apenas recentemente vem reaparecendo na literatura. O objetivo do quarto e último artigo que compõe essa tese é dar mais um passo na compreensão da justiça como ligada a um sentimento, concentrando-se no papel que o conceito de merecimento desempenha na intuição popular sobre as práticas de justiça

distributiva. A fim de aperfeiçoar o entendimento deste conceito, eu desenhei um experimento que visa a lançar luz sobre a relevância que as pessoas atribuem ao papel do merecimento na determinação da distribuição de renda entre os indivíduos na sociedade.

Como discutido anteriormente, a filosofia política contemporânea escolheu um caminho racionalista, tentando eliminar tudo aquilo relacionado à emoção da sua fundação racional para a moralidade humana. No entanto, os resultados das ciências empíricas, como mencionado, apontam precisamente para a estrada não trilhada: nossos julgamentos morais são gerados por um processo que envolve sim as nossas emoções.

Como esperado, uma das principais tentativas recentes de incorporação do reino do empírico nas teorias de justiça é também uma das abordagens que começa a reconhecer os aspectos emocionais da filosofia política e, consequentemente, a importância da ideia de merecimento. Em *Princípios de Justiça Social*, David Miller se posiciona ao lado de Hume na interpretação dos julgamentos de merecimento como intrinsecamente dependentes dos sentimentos de admiração e gratidão. Nas suas palavras,

> Se considerarmos as atitudes de admiração, aprovação, etc., fica claro que não as adotamos apenas como resposta àquelas qualidades que acreditamos terem sido voluntariamente adquiridas. Quando admiramos a habilidade de um músico, não perguntamos sobre a conduta que levou à sua aquisição antes de conceder a nossa admiração. A atitude é resultado direto da qualidade, uma vez que agora existe, e a pergunta, 'voluntariamente adquirido ou não?' simplesmente não é considerada. Se a estreita relação entre a avaliação de atitudes e merecimento é admitida, parece inconcebível que tais julgamentos como 'Green (o músico) merece reconhecimento' não devem ser feitos na mesma base: na base da habilidade apenas, sem referência à forma de sua aquisição. Essa é de fato a nossa prática. (Miller, 1976, p. 96; tradução própria)

É uma questão central da filosofia política se as intuições dos indivíduos sobre a justiça englobam ou não um princípio de merecimento. Na sequência dos trabalhos de Rawls, os liberais igualitários fizeram reivindicações de responsabilidade – e, consequentemente, de merecimento – praticamente desaparecer do debate sobre a justiça. Eles argumentam que a maior parte da nossa renda e da nossa riqueza é resultado do que se convencionou chamar de *sorte bruta*, e que o reconhecimento da

veracidade desse argumento é suficiente para demonstrar que princípios de merecimento não devem desempenhar nenhum papel na determinação da distribuição de renda entre os indivíduos. Em outras palavras, eles argumentam que a sorte bruta é suficiente para anular reivindicações de merecimento.

No entanto, existem razões para duvidar que essa visão seja de fato compartilhada pela maior parte das pessoas. Além disso, temos razões também para duvidar que os próprios filósofos tenham direito a esse ponto de vista. Em relação ao primeiro ponto, há um extenso corpo de pesquisa empírica que mostra que alegações sobre merecimento e responsabilidade constituem uma parte importante do conceito comum de justiça distributiva (Miller, 2003, Capítulo IV). Em relação ao segundo ponto, filósofos políticos, tais como David Miller, David Schmidtz e George Sher já começaram a responder negativamente à seguinte questão: reivindicações de sorte bruta realmente anulam reivindicações de merecimento?

Seguindo Hume, esses filósofos apelam para a indiferença do senso comum com relação às condições através das quais as bases do merecimento foram adquiridas. Nesse sentido, se a aceitação do princípio de merecimento repousa sobre o senso comum, é imperativo confirmar se essa é de fato a visão compartilhada pela população em geral. Apesar da extensa pesquisa empírica apresentada por diferentes cientistas sociais sobre o conceito de merecimento, não há evidências suficientes sobre diversas nuances desse conceito – principalmente no que diz respeito às intuições das pessoas sobre o papel da sorte bruta na distribuição de recursos.

Como resultado, várias perguntas permanecem sem resposta. Por exemplo, as pessoas realmente acreditam que a sorte bruta não anula reivindicações de merecimento, tal como os filósofos David Schmidtz e David Miller têm sugerido? Existem diferenças nessa crença de acordo com diferentes tipos de base de merecimento, tais como esforço, talento artístico, talento atlético, etc.?

Em um esforço para contribuir para este programa de pesquisa localizado na interseção da filosofia política e da psicologia política, Freiman & Nichols (2010) desenvolveram um experimento para esclarecer o seguinte conflito: a tendência observada de se atribuir "julgamentos de merecimento a indivíduos devido ao seu desempenho como um todo e, concomitantemente, restringir tais juízos apenas aos resultados que não podem ser atribuídos à sorte" (Freiman & Nichols, 2010, p.2; tradução própria). A hipótese dos autores é que este conflito se baseia na assimetria estabelecida na literatura experimental entre julgamentos realizados sob um contexto

abstrato e julgamentos realizados sobre condições concretas. Com base nessa hipótese, eles buscam mostrar que "indivíduos defrontados com uma pergunta puramente abstrata sobre merecimento são mais propensos a dar respostas em conformidade com a restrição de sorte bruta do que indivíduos defrontados com um caso concreto sobre um indivíduo em particular" (Freiman & Nichols, 2010, p.2; tradução própria). Os resultados empíricos encontrados pelos autores corroboram a sua hipótese. Entretanto, existem alguns problemas com o desenho do seu experimento que colocam em dúvida a validade desses resultados.

Nesse contexto, o objetivo do experimento que eu apresento no quarto artigo que forma essa tese é duplo: (i) aperfeiçoar o design experimental usado por Freiman & Nichols (2010), corrigindo seus problemas; e (ii) fornecer dados adicionais sobre as nuances do conceito comum de merecimento. O primeiro objetivo baseia-se na premissa de que os achados de Freiman & Nichols (2010) foram resultado de um erro metodológico na formulação do cenário abstrato. Dessa forma, eu elaborei novos casos abstratos que corrigem esse erro, de modo a testar se a hipótese inicial se mantém sob o design experimental revisto.

O segundo objetivo é explorar algumas características do conceito de merecimento que são ignoradas em seu trabalho. Freiman & Nichols utilizam em seu experimento apenas três cenários: um abstrato e dois concretos. Como resultado deste número limitado de cenários, eles não são capazes de explorar uma ampla gama de intuições das pessoas sobre merecimento. Por exemplo, eles não são capazes de explicitar características relevantes para um melhor entendimento do nosso uso ordinário desse conceito, tais como: como a base de merecimento foi gerada – foi o resultado de *sorte natural* ou de *sorte social?* Nesse sentido, o segundo objetivo é aprimorar o design do experimento através da criação de novos cenários capazes de iluminar essas e outras características do nosso conceito compartilhado de merecimento.

***

***

Assim, a presente tese é constituída pela defesa de uma filosofia política empiricamente informada de forma mais substantiva e, concomitantemente, por um exercício de desenvolvimento de duas possíveis implicações de tal mudança metodológica.

Daniela Goya Tocchetto
Porto Alegre, RS, Brasil
Março de 2014

List of Contents

List of Figures

Introduction

Contemporary political philosophers are broadly concerned with the study of human social organization. More specifically, they aim at the elaboration of a set of principles capable of stating how we should organize our lives not as atomistic individual beings, but as active members of cooperative endeavors.[2] How should we understand our mutual responsibilities to one another as members of a society? What sorts of treatments do we rightly owe each other?

In the search for the principles that will provide the answers to the above and related questions, the main virtue in which contemporary political philosophers are interested is the virtue of *Justice*–according to Rawls (1971), the primary virtue of social institutions. Within the realm of justice, *Distributive Justice* emerges as one of the central areas of research in political philosophy today.

Principles of distributive justice are meant to guide the workings of social institutions with respect to the allocation of burdens and advantages among the members of a society, such as the allocation of education, medical treatment, and taxes. In this way, a theory of distributive justice is crucial to the development of a fair and well-functioning society–even if this theory is solely tacit and has not been explicitly developed, as in small ancient societies. Moreover, in order to be successful, a theory of justice must be able to "persuade people to regulate their intuitive sense of justice by its principles and allow this hope to be realized" (Miller, p.21).

In spite of the central role played by theories of distributive justice in contemporary political philosophy, there is to date nothing close to a consensus on which set of principles should guide the allocation of social and economic benefits and burdens amongst individuals. All the more disturbing, as nicely highlighted by David Miller:

[2] It is worth noting here that some philosophers envision these principles as entailing an atomistic view of society.

> (…) contemporary liberal moral and political philosophy presents a spectacle of continuing deep disagreement between rival theories of justice. Each theory claims to embody demonstrable truth, but there is no reason to think that the contest between them will ever be resolved. (2003, p.112)

In the face of the practical relevance of the subject and the current state of comprehensive dissent about which sort of principles of justice we should abide to, it seems imperative to address the following question: why do we find ourselves in this present state of "deep disagreement"? There is no doubt that a great part of the answer is related to the complexity of the matter–justice is indeed not straightforward! Yet attributing the problem solely to its subject complexity would be an acknowledgment of inevitable failure, an acceptance of the impossibility of the pursuit. Therein rests the necessity of exploring an alternative explanation for this worrisome state of affairs.

In this dissertation, I am going to argue that a fruitful alternative explanation can be encountered in the investigation of a notable feature shared by most contemporary political philosophical theories, namely, their methodological approach. Contemporary political philosophers rather frequently rely in the method of so-called armchair philosophy, which is characterized by the process of abstraction from intuitions. The idea is to start from what philosophers *claim* to constitute people's basic intuitions and, from there, build a rationally coherent set of abstract principles. In this manner, philosophers can do without empirical information about human morality. Empiricists such as Hume have already pointed out the flaws of this methodology, but the majority of political philosophers seem to have currently sided with Kant on which is the proper way of developing first order normative theories.

This methodological choice has made contemporary political philosophers reluctant to more fully address the results that have been emerging from the empirical sciences–especially those results from the natural sciences. This reluctance would be appropriate were there consistent reasons for the denial of empirical data as an important resource for normative theorizing. Nonetheless the reasons that have been presented by philosophers are not consistent. Quite the contrary, there are

surmounting reasons for the opposite methodological stance[3] –namely, seriously considering all *relevant empirical evidence*[4] in normative theorizing.

This hesitancy is not only inappropriate, but it is also responsible for bringing about a number of distressing consequences. Most notably, not properly addressing the recent results from empirical research on moral behavior has made philosophers oblivious to the sentimental nature of human morality. This is a serious consequence that may not only lead to a dangerous disconnection between theories of justice and human *actual* moral behavior, but could also be preventing political philosophers from developing alternative theories that may well help to minimize the aforementioned comprehensive dissent in matters of justice.

The Reasons for a broader embracement of empirical evidence

The relevance of empirical findings concerning human morality leads us to an examination of the arguments that contemporary political philosophers have historically relied on so as to overlook a wide array of empirical data in normative theorizing.[5] According to David Miller, contemporary political philosophers have generally appealed to two main arguments in order to refrain from getting their *hands empirically dirty* (2003, p.42). The first argument states that empirical research is unable to reveal people's *considered judgments* about justice, while the second argument relies on the logical gap between what people's actual beliefs *are* and what they *should* be. This second argument amounts to the widely known logical impossibility of deriving 'ought' statements from 'is' statements–the so-called natural fallacy. Hence Miller (2003) claims that contemporary political philosophers' reluctance to give empirical evidence a more significant role in the development of first-order principles of justice derives primarily from a distinction between justification and acceptance: showing that a belief is accepted, philosophers assert, neither shows that it is justified nor that it is normatively obligatory.

On the former, the 'considered judgments' critique regards the folk's lack of specialized knowledge on morality. It is not a claim about the irrelevance of folk

[3] These reasons are the focus of the first paper that composes this dissertation.
[4] Whenever I mention *relevant* empirical evidence, I refer to all evidence about human moral intuitions, beliefs, and behavior.
[5] These arguments are the object of the first paper in this dissertation.

intuition for normative theorizing, which would be rather strange in face of the long tradition of reliance on human intuitions in moral and political philosophy. For instance, philosophers as distinct as Aristotle and Rawls explicitly appeal to folk intuitions about justice in the development of their respective theories. Thus what keeps the political philosopher from more heavily relying on empirical results is a methodological stance: a claim that to reach human moral intuitions from the armchair is the appropriate philosophical way of proceeding given the incapacity of the general population to properly formulate its considered judgments about morality. Numerous contemporary political philosophers adopt this same methodological attitude.

On the latter, the 'natural fallacy' critique regards the logical conditions limiting or allowing the collaboration of normative philosophical theories and empirical sciences; it is strictly a logical claim. Hume's Law, as stated in *A Treatise of Human Nature*, affirms that while the logical value of being true or false can be attached to empirical statements, this is not possible for normative statements. Thus it is logically inadmissible to infer statements of *ought* from statements of *is*. The issue at stake here, I will argue, is that one does *not* have to deny this logical impossibility in order to embrace an empirically informed political philosophy.

It is important to stress that advocating for a broader empirical understanding of the main concept of political philosophy–namely, justice–implicates neither a naive endorsement of accepted beliefs as justified ones, nor an infringement of logical rules. Instead, the recognition of the relevance of empirical data merely constitutes an acknowledgment of its proper role in helping to develop political theories that are both reliable and feasible, as argued in the first paper that composes this dissertation.

The arguments in favor of taking empirical evidence seriously when theorizing about justice are thoroughly examined in this first paper. They are organized, for the purposes of addressing the main contentions presented by the so-called "purist" political philosophers, into two main groups: (i) against the 'considered judgments' critique; and (ii) against the 'natural fallacy' critique. This second group of arguments does not imply a refutation of the natural fallacy; its logical claim remains valid. The arguments that are exposed in this subsection are only intended to refute the use of the natural fallacy as an impediment to interdisciplinary research in political philosophy.

After examining all the arguments, it will be clear that we have an over-determining assemblage of reasons for embracing the incorporation of all relevant empirical evidence into political philosophical theorizing. This incorporation can take place in three distinct levels of the process of theorizing about justice, namely: (i) the meta-philosophical level, (ii) the first-order philosophical level, and (iii) the applied ethics level. Contemporary political philosophers have already been drawing on empirical data in the applied ethics level, yet there is still a lot of space for bringing this data into play in the remaining two levels.

Just to mention one of the reasons in support of an empirically informed political philosophy, consider, for instance, the lack of practical guidance provided by principles of justice that are dissociated from real patterns of human behavior. There are several other good reasons for political philosophers to more carefully consider empirical evidence regarding ethical human behavior, and they are given the proper attention in the above-mentioned first paper.

The Rational Character of Contemporary Theories of Justice

The most obvious consequence of the methodological approach chosen by contemporary political philosophers is the current prevalence of a rationalist trend in theories of justice. In order to provide a more specific example of this trend, I will focus my attention on the narrower subset of theories of distributive justice. We can identify several contemporary theories of distributive justice. These theories vary across the many dimensions that comprise distributive principles, such as: (i) what is relevant–income, wealth, opportunities, jobs, welfare, utility, etc.; (ii) the nature of the recipients–individuals, groups, classes, etc.; and (iii) the distributive rule–equality, maximization, according to individual characteristics, according to free transactions, etc. The take home lesson is that, in despite of all this diversity, the majority of these theories remain constant along one fundamental dimension: rationalism.

In this context, we can roughly divide the main contemporary theories of distributive justice into two broad groups: (i) Libertarianism, and (ii) Liberal Egalitarianism. On the one hand, libertarians are solely concerned with the protection of individual rights, firstly envisaged by Locke as the natural rights to life, liberty and property. On the other hand, liberal egalitarians include all political philosophers who

share with the libertarians the embracement of the intrinsic value of autonomy and the consequent relevance of individual liberties, while at the same time acknowledging the injustices engendered by the discrepancies in human conditions due to the effects of the social and the natural lotteries.

In addition to rationalism, the aforementioned theories share another remarkable feature, which is also a consequence of contemporary political philosophers' methodological choice: their rationalist conception of human morality is miles away from what the empirical sciences have been revealing about our moral nature. Hence an empirically informed political philosophy may trigger a significant change in our understanding of justice.

Primatologists, evolutionary biologists, moral and social psychologists, and neuroscientists–among other scientists–have in the past three decades gathered significant data indicating that our moral rules are more emotional than the rationalist crowd has suggested and anticipated. The literature reports extensive evidence that our moral judgments are brought about by emotional reactions, and that we are easily morally dumbfounded by our own moral intuitions (Haidt et al., 1993). Numerous other experiments have shown that our moral judgments are strongly affected by environmental cues, heuristics and biases, emotional intuitions, and the like (e.g. Cushman et al., 2006; Greene et al., 2004; Sinnott-Armstrong et al., 2010; Wheatley & Haidt, 2005). And this is just to mention one stream of all the empirical data that urges to be properly addressed by contemporary political philosophers.

Thus the aim of the second paper that composes this dissertation is to provide an extensive review of the existing empirical literature on human intuitions, beliefs, and behaviors related to the concepts of justice and fairness. This review includes some of the most significant research involving these concepts during the past three decades in the areas of primatology, evolutionary biology, behavioral economics, moral psychology, political and social psychology, and neuroscience. The goal of this second paper is twofold: (i) to make all these novel research programs and some of its interesting results easily available for political philosophers; and (ii) in so doing, to fuel the development of a more fully empirically informed political philosophy, an area that is by nature multidisciplinary and should therefore be treated as such.

The Implications of Properly Addressing the Relevant Empirical Evidence

After appraising the arguments for and against the role of the empirical sciences in contemporary political philosophy and exploring the recent results from empirical research on moral behavior, I move on to discuss some implications of this debate. Firstly, I argue that taking empirical evidence seriously will trigger the embracement of a more sentimentalist political philosophy. Secondly, I argue that empirical evidence will also have the consequence of altering the status we grant to principles of desert in contemporary theories of justice.

(i) Sentimentalism

As previously discussed, contemporary political philosophers have relegated affective states to a secondary role in their theories of justice. Of the two enlightenments that occurred in the eighteenth century, contemporary political philosophers–having Rawls as their main mentor–have widely embraced the rationalist one (Frazer, 2010). A paradigmatic example of this secondary role assigned to emotions in current political philosophy is the fact that Rawls himself only analyzed our affective structure after having already constructed a solid rational basis for both his principles. In this sense, Rawls claims that our emotions only play a role as either proving to be fit or to be an obstacle to the application of the principles of justice.

Yet, as we have already briefly discussed, empirical scientists have in the past decades provided surmounting evidence suggesting that our moral rules are less Kantian than the rationalist crowd could have assumed. It is at the very least surprising that, despite all the evidence pointing towards an emotional account of morality, contemporary political philosophers have in their majority remained alien to moral sentimentalism.[6] In this context, the focus of the third paper that constitutes this dissertation is to make the case for a sentimentalist turn in contemporary political philosophy.

In order to do so I will argue, along with Frazer (2010), that the main reasons presented by political philosophers for the dismissal of moral sentimentalism cannot

[6] This is not to say that political philosophers have ignored our affective states as completely irrelevant to justice; it is instead a claim that they have only acknowledged them insofar as they constitute an important step in the judgment of the stability of institutional arrangements–and this acknowledgment will be shown to be insufficient.

be sustained after due scrutiny. For instance, two of the main arguments that have been presented against moral sentimentalism are (i) the fear of falling into a descriptive account of morality, with no normative power, and (ii) the problem of the separateness of persons, as pointed out by Rawls. In this second critique, Rawls claims that our experience of sympathy leads us to overlook the inviolability of individuals. If sympathy did in fact blur the distinctions between us, reflective sentimentalism would indeed be incompatible with a liberal theory of justice built around individual rights and the inviolability of distinct persons (Frazer, 2010). Yet neither the first nor the second threats are of real danger to moral sentimentalists, as will be properly demonstrated–along with the refutation of two additional important critiques of moral sentimentalism.

In his recent book, *The Enlightenment of Sympathy*, Michael Frazer attempts to begin the hard work of building a more sentimentalist view of justice. Along similar lines, the third paper that composes this dissertation also constitutes such an attempt. In the final section of the paper I therefore (rather preliminarily) discuss some of the implications of a sentimentalist turn in political philosophy. To cite one important implication of such sentimentalist turn, consider how contemporary political philosophers have advocated for a political debate grounded solely on rational argumentation. Emotions are usually looked down on as argumentative resources in political deliberation. Yet if moral sentimentalism is right, this is a distressing state of affairs. This reason-based tendency has triggered an emotional disengagement in the political scenery, which may in turn have helped to make our society even more individualistic and less concerned with the general welfare. Perhaps if we start assigning to rhetoric a larger role in the political sphere we will be able to more properly engage with our moral emotions and, as a consequence, generate greater social cohesion.

(ii) Desert

The concept of desert has largely disappeared from contemporary political philosophy in the wake of Rawls's work and has been only recently reappearing in the literature–especially in the still incipient sentimentalist renascence of justice. In this context, the aim of the fourth paper is to take a further step in the comprehension of justice in a more sentimentalist fashion by focusing on the role that desert plays in the folk intuition concerning practices of distributive justice. In order to further refine the

understanding of this role, I designed an experiment to shed light on the intricate relation between desert and luck in the distribution of income across individuals in society.

As previously discussed, contemporary political philosophy has taken the rationalist road, attempting to eliminate all that is related to affect and emotion from its supposedly solid rational foundation. Yet, as we have discussed, results from the empirical sciences point precisely to the road not taken: our moral judgments are engrained with affect. Unsurprisingly, one of the leading contemporary attempts to incorporate empirical evidence in theories of distributive justice is also one that begins to recognize the emotional aspects of our morality–and, along with these aspects, the role of principles of desert. In *Principles of Social Justice*, David Miller sides with Hume in interpreting judgments of desert as intrinsically dependent on feelings of admiration and gratitude. As he writes:

> If we consider the attitudes of admiration, approval, etc., it is plain that we do not adopt them only towards qualities believed to be voluntarily acquired. When we admire the superlative skill of a musician, we do not ask about the conduct which led to its acquisition before granting our admiration. The attitude is held directly towards the quality as it now exists, and the question, 'voluntarily acquired or not?' is simply not considered. If the close relation between appraising attitudes and desert is admitted, it seems inconceivable that such judgments as 'Green (the musician) deserves recognition' should not be made on the same basis: on the basis of the skill alone, without reference to the manner of its acquisition. And this is indeed our practice. (Miller, 1976, p.96)

It is a major question in political philosophy whether or not individuals' intuitions about justice encompass the principle of desert. Following the work of Rawls, liberal egalitarians made claims of responsibility–and consequently, desert–practically disappear from the justice scene. They argue that most–if not all–of our income and wealth comes from brute luck, and that this fact alone is sufficient to show that desert should play no role in determining the distribution of income amongst individuals. In other words, they argue that claims of brute luck are sufficient to nullify claims of desert.

Yet we have reason to doubt that this view is shared by the folk. Moreover, we have reason to doubt that philosophers themselves are entitled to this view. Regarding the former doubt, there is an extensive body of empirical research showing that claims

about desert and responsibility constitute an important part of the folk's concept of distributive justice (Miller, 2003, Chapter Four); regarding the latter doubt, political philosophers such as David Miller, David Schmidtz, and George Sher have begun to pose the question: do claims of brute luck really nullify claims of desert?

Following Hume, these philosophers appeal to commonsense morality's indifference to the conditions under which desert bases are acquired. Yet if the embracement of desert should rest on commonsense morality, it is imperative to confirm if such is indeed the folk's view. Despite the fair amount of evidence collected by a range of different social scientists on the folk's concept of justice, there is not enough evidence on the nuances of the concept of desert.

As a result, several unanswered empirical questions remain. For instance, do the folk actually believe that brute luck does not nullify claims of desert (as the aforementioned researchers have suggested)? Are there differences in this belief according to different kinds of desert basis–effort, artistic talent, athletic talent, etc.? Are there differences in desert beliefs according to the kind of desert; for instance, economic or moral appraisal?

In an effort to contribute to this research program at the cross roads of political philosophy and political psychology, Freiman & Nichols (2010) designed an experiment to shed light on the following conflict: the tendencies observed among the folk to at the same time "judge individuals' deserts in terms of their performance alone and to restrict such judgments to those products within their control" (Freiman & Nichols, 2010, p.2). Their idea is that this conflict rests on the established asymmetry between judgments made either under abstract or under concrete conditions, and their hypothesis is that "subjects presented with a purely abstract question about desert would be more likely to give responses conforming to the brute luck constraint than subjects presented with a concrete case about a particular individual" (Freiman & Nichols, 2010, p.2). While their findings appear to support their prediction, there are some issues with their experimental design that I seek to investigate (and avoid) with my present research.

Thus the goal of the experiment presented in the fourth paper that forms this dissertation is twofold: (i) to improve upon the experimental design used by Freiman & Nichols (2010); and (ii) to provide additional data on the nuances of the folk's concept of desert. The first goal rests on the premise that the findings reported in Freiman & Nichols (2010) were driven by a misformulation of the abstract scenario.

They only used one abstract case, which involved agents who benefit from genetic advantages. However, Friedman & Nichols did not distinguish between different types of genetic advantages. Moreover, they did not specify which kind of genetic advantage was conducive to the individual's higher level of income. This under specification failed to control for alternative interpretations of their case: participants were unwillingly invited to fill in the details not explicit in the case. Therefore, I designed new abstract cases that addressed this misformulation. My results revealed that their working hypothesis did no longer hold under the revised experimental design.

The second and related goal is to explore some features of the concept of desert that are ignored in their work. Friedman & Nichols used very few scenarios: only one abstract and two concrete. As a result of this limited number of cases they were unable to explore a wide range of people's intuitions about desert. For instance, they were not able to address individual's intuitions on different sources of brute luck, namely, natural or social luck. Hence in this fourth paper I have built on their experiment, providing new scenarios that explored some of these under investigated features.

***

In a nutshell, the present dissertation constitutes a defense of an empirically informed political philosophy and, at the same time, an exercise in the preliminary development of two crucial implications of such empirically informed practice.

The Role of the Empirical Sciences in Political Philosophy

(…) we should always carefully separate the empirical from the rational part, and prefix to Physics proper (or empirical physics) a metaphysic of nature, and to practical anthropology a metaphysic of morals, which must be carefully cleared of everything empirical, so that we may know how much can be accomplished by pure reason in both cases.
(Kant, Groundwork)

Now it is only a pure philosophy that we can look for the moral law in its purity and genuineness (and, in a practical matter, this is of the utmost consequence): we must, therefore, begin with pure philosophy (metaphysic), and without it there cannot be any moral philosophy at all. That which mingles these pure principles with the empirical does not deserve the name of philosophy (for what distinguishes philosophy from common rational knowledge is that it treats in separate sciences what the latter only comprehends confusedly); much less does it deserve that of moral philosophy, since by this confusion it even spoils the purity of morals themselves, and counteracts its own end.
(Kant, Groundwork)

Introduction

The above quotes by Immanuel Kant vividly instantiate a methodological position that has been prevalent in contemporary[7] political philosophy at least since the publication of John Rawls' groundbreaking *A Theory of Justice*, in 1971–namely, the adherence to ideal theories and rationalism as the proper way to arrive at principles of justice. This methodological stance has been conducive to the present state of affairs in political philosophy, characterized by an ongoing rationalist debate easily recognized in the endless contemporary publications in the major journals of the field. As a consequence of this idealist and rationalist attitude, contemporary political philosophers have been making very little use of surmounting evidence about human morality gathered by primatologists, evolutionary biologists, psychologists, experimental economists, and neuroscientists.

There are certainly remarkable exceptions. Even neo-Kantian political philosophers such as Rawls himself have been sensitive to empirical findings from a

[7] "(…) what is connoted by our focus on *contemporary* political philosophy? Within the analytical tradition of thought, as that affects both philosophy and other disciplines, political philosophy has become an active and central area of research in the past three or four decades; it had enjoyed a similar status in the nineteenth century but had slipped to the margins for much of the twentieth. In directing the *Companion* to contemporary political philosophy, we mean to focus on this recent work" (*Companion*, 2012, p. xvii).

subset of fields, like economics and other social sciences.[8] In Rawls's case, the degree to which he demonstrated being sensitive to the workings of the empirical world is especially noteworthy. Most political philosophers address empirical data about human behavior solely after the principles of justice are in place, so as to check the *feasibility* and the *stability* of their proposed set of justice principles. Rawls does in part fit with this general way of proceeding shared by the majority of his fellow political philosophers. Nonetheless, he goes beyond this standard modus operandi.

In the second part of his second principle of justice, he makes a concession to unequal distributions of income insofar as this inequality is capable of improving the lives of those least advantaged in society–the so-called *difference principle*. This concession is the result of incorporating the teachings of economics, more specifically, the idea that incentives are necessary in order for people to perform their best. In this manner, Rawls fully acknowledges and addresses the empirically demonstrated tradeoff between efficiency and equality, shaping the form of his second principle of justice so as to properly incorporate this economic fact.

Hence my claim in the present paper does not amount to stating that contemporary political philosophers have been completely oblivious to the results of the empirical sciences. Neither does it amount to a plea for dismissing the current methodology, replacing it for some entirely new way of conducing the development of theories of justice. My argument rests on the identification of three problems with the manner in which political philosophers have assimilated the relevance of empirical evidence. Firstly, political philosophers have not yet embraced all sorts of empirical evidence–this is especially true in relation to the findings from the natural sciences. Secondly, the degree to which philosophers have taken account of the results from empirical sciences is still rather incipient. In this sense, it seems necessary to give all the relevant empirical evidence, from the social and the natural sciences, due consideration. Thirdly, most political philosophers have incorporated empirical findings late in the process of the development of their theories of justice. That is, they have turned their attention to actual human moral behavior *after* the principles of justice are already in place, solely in order to check whether these principles pass the tests of being both feasible and stable.

[8] It is important to stress at this point that political philosophers have more easily incorporated empirical findings from the social sciences than from the natural sciences.

Rawls is an example of a philosopher that incorporated empirical findings in the formulation of his principles. Nonetheless, he still falls short of addressing a variety of empirical sciences that are relevant to the understanding of human moral behavior. All the more, Rawls remains a Kantian in his ultimate foundation of justice, and the rationalist flavor that underlies his works is quite explicit. This is most likely a consequence of his methodological choice to construct an *ideal* theory of justice. As has been the focus of very recent debate, perhaps we should pay more attention the to development of *nonideal* theories of justice.[9]

In his book *The Enlightenment of Sympathy*, Michael Frazer offers an interesting analysis of the historical reasons that lead contemporary political philosophers to be hesitant about embracing a broader incorporation of empirical evidence in their theories of justice. Frazer (2010) claims that in the eighteenth century we have witnessed the emergence of two distinct enlightenments: the Kantian enlightenment of reason, and the Humean enlightenment of sentiment. Contemporary political philosophers, Frazer alleges, followed Kant down the rationalist path. As a consequence, they embraced our rational faculties alone as the proper ground for all normative systems. Once it is agreed that human morality can be grounded solely in our rational faculty alone, all moral systems are understood as axiomatic systems based on ideas such as inalienable rights and duties, and usually guided by the core value attached to human dignity. And, once we start working under an axiomatic framework, empirical evidence becomes less and less useful–this is my hypothesis for why we stand where we presently stand.

Recently, however, some exceptions to this traditional approach have been emerging. Pluralists such as David Miller and Michael Walzer have developed theories of justice that heavily rely on folk intuitions about justice–thus paying closer attention to findings from the empirical sciences regarding human moral beliefs and behavior. In the development of their respective political theories they demonstrate a high concern with pragmatic viability, which results in the endorsement of a plurality of principles amongst the different spheres of human life. The observance of this

[9] I refer here to nonideal theories as exposed by Amartya Sen in his latest book, *The Idea of Justice*. In this book, he argues against the development of what he calls transcendental theories of justice, and in favor of the so-called comparative approach to justice theories. Yet this debate is not the focus of this paper. Before one argues for the superiority of nonideal or comparative theories, one has to be certain that all arguments against the broader incorporation of empirical evidence in the development of political philosophical theories of justice are not valid. Therefore the present focus in the assessment of these arguments.

plurality emancipates philosophers from the supposed necessity of having a single set of principles that is valid for all types of social contexts.

In a distinct vein, utilitarian theorists such as Peter Singer and Peter Unger are also more open to the usage of empirical evidence whenever it is relevant to their subject matter. The difference is that utilitarians draw on the empirical sciences not to better understand folk intuition, but to downplay its authority on the derivation of moral principles and, subsequently, to argue for consequentialism. Notwithstanding these recent efforts, empirically informed theorizing about justice is far from being part of mainstream contemporary political philosophy.

Hence the fact that the most prominent political philosopher of the twentieth century incorporates a limited range of empirical evidence in his theorizing does not offset the pressing need for a more significant role for the empirical sciences in the development of contemporary political philosophical theories of justice. In this context, we should ask ourselves: are there any reasons to stand in opposition to the aforementioned empirical scarcity identified in the contemporary political philosophical literature on justice? After all, why would all sorts of empirical evidence about human moral behavior be significant for Political Philosophy? And if so, in which ways would it be significant? To address these questions is the aim of the present paper.

The paper is structured in five sections. The second section, following this introduction, presents the two main arguments philosophers have maintained against a broader empirically informed political philosophy. Subsequently, the third section addresses these critiques, expressing all the arguments for their dismissal. The fourth section brings out a more ambitious argument in favor of the relevance of empirical data for theorizing about justice. Finally, the fifth section provides a brief discussion about the proper role of the empirical sciences in contemporary political philosophy.

2. The Arguments Against an Empirically Informed Philosophy

According to David Miller, philosophers generally appeal to two main arguments in order to refrain from getting their hands 'empirically dirty' (2003, p. 42). The first argument states that empirical research is unable to reveal people's *considered judgments* about justice; while the second argument relies on the logical gap between what people's actual beliefs *are* and what they *should* be–the widely

known argument for the logical impossibility of deriving an *ought* from an *is* (so-called 'natural fallacy'). In this sense, Miller (2003) claims that political philosophers' reluctance to give all relevant empirical evidence a significant role to play in the development of justice theories derives primarily from a distinction between justification and acceptance. In this sense, showing that a belief is accepted, philosophers assert, shows neither that it is justified nor that it is normatively obligatory.

The 'considered judgments' critique addresses the folk's alleged lack of specialized knowledge about morality. It is not a claim about the irrelevance of folk intuition for moral theorizing; a point that would be at the very least strange in face of the long tradition of reliance on human intuitions in moral and political philosophy. For instance, philosophers as distinct as Aristotle and Rawls explicitly appeal to folk intuitions about justice in the development of their respective theories. Thus what prevents the philosopher from relying on the empirical sciences is a methodological stance–a claim that armchair theorizing is the appropriate philosophical way of proceeding given the incapacity of the general population to properly formulate its considered judgments about morality. Numerous philosophers adopt the same methodological attitude. Their claim is not that intuitions are irrelevant; it is specifically that *folk* intuitions are irrelevant.

The 'natural fallacy' critique addresses the logical conditions limiting or allowing the collaboration of normative ethical theories and empirical sciences. That is, it constitutes a logical claim. Hume states in *A Treatise of Human Nature* that, while the logical value of being true or false can be attached to empirical statements, the same is not possible for normative statements. Thus, it is logically inadmissible to infer *ought* statements from *is* statements. The issue at stake here, I will argue, is that one does not have to deny this logical impossibility in order to embrace a thoroughly empirically informed political philosophy.

It is important to stress that advocating for the importance of a broad empirical understanding of the main concept of political philosophy, namely justice, neither implicates a naive endorsement of accepted beliefs as justified ones, nor constitutes an infringement of logical rules. The recognition of the relevance of empirical data about the nature of human morality constitutes an acknowledgment of its proper role in helping to develop political theories that are both reliable and feasible, as will be argued in the remainder of this paper.

3. The Arguments in Favor of an Empirically Informed Philosophy

In this section I will address the main arguments against an active collaboration between political philosophy and all of the relevant empirical sciences. This collaboration could be understood in two different ways: political theory informing the empirical sciences, and the empirical sciences informing political theory. Here I restrict my attention to the latter form of collaboration.

In order to address the main contentions against a substantially empirically informed political philosophy, as described in the previous section, the arguments in favor of taking the relevant empirical evidence seriously will be organized under two broad groups: (i) Against the 'Considered Judgments Critique,' and (ii) Against the 'Natural Fallacy Critique.'

(i) Against the 'Considered Judgments Critique'

There are over-determining reasons for the dismissal of this critique. For starters, it is crucial to bear in mind that this critique is not aimed at the dependence on human intuitions in political philosophical theorizing. The central point of the 'considered judgments critique' is the alleged inappropriateness of employing the methods of the empirical sciences to arrive at philosophical intuitions. The mainstream method that philosophers draw on to assess our intuitions is the so-called armchair philosophical method. This method is characterized by the lonely reasoning about the issues at stake, so as to enable the philosopher to envisage by introspection alone which intuitions are relevant to his subject of interest.

In this respect, there are two main grounds on which we should be suspicious of armchair philosophy and in favor of an empirically informed practice. Firstly–and unsurprisingly, there is widespread disagreement among philosophers about which moral intuitions are universally shared by the laypersons. The pervasiveness of this disagreement is a sign that there is something suspicious about armchair philosophy. After all, how can an accurate method of arriving at our *shared* considered judgments result in the attainment of distinct–and even divergent–claims? This first argument does not constitute a claim that empirical methods are more adequate for normative philosophical theorizing; it is intended to undermine the superiority of armchair philosophy. I will call this first argument 'The Disagreement Argument.'

Secondly, the incoherence we encounter amongst folk intuitions does not itself constitute an impediment to the incorporation of commonsense morality in political philosophical theorizing. Many philosophers have argued against relying on folk intuitions based on the claim that it is of the utmost relevance to be able to arrive at a coherent set of intuitions and that, in order to do so, one needs philosophical specialized training. For those philosophers, the incoherence we encounter among the folk is an indication that they lack such specialized knowledge. I will argue against this claim. Moreover, I will contend that folk intuitions serve to illuminate political theories in several ways. This is a positive argument for the incorporation of empirical methods in political philosophy. This second argument is the so-called 'Expertise Defense Argument.'

The Disagreement Argument

One instance of the philosophical disagreements about folk intuitions generated via armchair methodology is provided by an analysis of the traditional 'justice as impartiality' approach (Frohlich & Oppenheimer, 1992). This conception of justice can be found in a variety of cultures and historical periods, as explicitly exemplified by the pervasiveness of the *Golden Rule*: "do unto others as you would have them do unto you." Contemporarily John Rawls and John Harsanyi, defenders of the opposite ethical systems liberal egalitarianism and utilitarianism, respectively, have both endorsed this methodology. Frohlich & Oppenheimer (1992) show that, from a theoretical perspective, they should both have arrived at precisely the same principles. Yet this is not what happened.

In order to establish their claim, Frohlich & Oppenheimer (1992) state the syllogism underlying the methodology employed by Rawls and Harsanyi as follows:

(i) C1,…,Cn are the ideal conditions of impartiality.
(ii) Any principle unanimously accepted under ideal conditions of impartiality is a valid principle of justice.
(iii) Under C1,…,Cn principle P would be accepted unanimously.
(iv) Therefore P is a valid principle of justice.

Rawls and Harsanyi provide an empirical-content free reasoning for the establishment of the principles that would be unanimously accepted under conditions

Cn, and end up arriving at diametrically opposed results. How to settle this disagreement? Here is where we can show that empirical evidence is able to play a crucial role. Instead of relying on assumptions from the armchair about which are the folk's considered judgments about justice under conditions of impartiality, Frohlich & Oppenheimer (1992) claim we should move to empirical research in order to actually discover the folk's judgments. In this sense, they provide a new syllogism, now with empirical content:

> (i empirical) C*1,…,C*n are experimental approximations to the ideal conditions of impartiality.
> (ii-a empirical) Any principle unanimously agreed on under experimental conditions C*1,…,C*n has a claim to be a valid principle of justice.
> (ii-b empirical) Any principle incapable of getting substantial support under experimental conditions C*1,…,C*n can be presumed to be rejectable as a valid principle of justice.
> (iii empirical) Under experimental conditions C*1,…,C*n principle P is unanimously agreed on and principle Q is incapable of getting substantial support.
> (iv empirical) Therefore P has a claim to be a valid principle of justice and Q can be presumed rejectable as valid principle of justice.

Frohlich & Oppenheimer (1992) contend that, once we achieve philosophical agreement on the role of impartial procedures for judgments of justice and on the ideal conditions of impartiality, the specific content of the principles will be better arrived at through the proper design of experiments than through armchair reasoning. That is, armchair philosophy would still have an essential function, namely, the definition of the appropriate role of impartiality and of its ideal conditions. As stated by the authors, "The deeper question really is whether the model of an impartial outside observer à la Smith or the model of an involved person under the veil of ignorance à la Rawls or Harsanyi is better suited for judgments on justice and injustice" (Frohlich & Oppenheimer, 1992, p. 64).

This example illustrates one possible relation between empirical inquiry and theories of justice, which has already generated an entire research program in political science: the use of laboratory experiments, usually with college students, designed to

reveal the judgments of individuals under controlled conditions of impartiality.[10] Unsurprisingly, the political philosophical community has largely ignored this research program.

In the face of widespread disagreement amongst philosophers about which intuitions are universally valid, we cannot help questioning the alleged superiority of the armchair methodology. After all, if it is assumed both that there are *common* intuitions *shared* by *every* average human being concerning moral issues and that the right method to arrive at these intuitions is armchair philosophy, how can one justify different philosophers arriving at incongruent intuitions via the exact same supposedly accurate and impartial method?

The Expertise Defense Argument

Another reason one can present in favor of an empirically informed methodology in political philosophy is the aforementioned claim that incoherence among folk intuitions is not sufficient for their dismissal as irrelevant. Quite the contrary, folk intuitions illuminate political philosophical theorizing in several ways.

One of the usual routes philosophers take to argue that laypersons' incoherent intuitions are a sign of their incapacity to achieve considered judgments is the so-called 'expertise defense.' The expertise defense maintains that philosopher's professional training is a necessary condition for the attainment of accurate philosophical intuitions (Feltz & Cokely, 2012, p. 238).

Yet this defense is not sustainable. There is overwhelming data showing that expert philosophers behave in much the same way as the laypersons. Moreover, there is data revealing that personality traits exert influence on the intuitions of verifiable experts–and that they also remain unaware of this influence.[11]

Additionally, contemporary research in moral psychology has been providing cumulative evidence that most human intuitions are messier than we had anticipated–and the philosopher's intuitions are not immune to this messiness. Yet the fact that our intuitions are muddled does not straightforwardly imply that we should not take them seriously, or that the task of the philosopher is to render them coherent–as one could flippantly think.

---

[10] For a good review, see 'Empirical Social Choice: Questionnaire-Experimental Studies on Distributive Justice,' Gaertner & Schokkaert, 2010.

[11] For a review of this evidence, see Feltz & Cokely (2012).

On the contrary, this *incoherence* may be illuminating in several ways–many of which we may to date be still unaware. There are at least four ways in which data about this incoherence has already been illuminating present research: (i) uncovering the biases that generate some of these incoherencies; (ii) revealing the distinctive psychological and neurological mechanisms responsible for the generation of these differences in our moral intuitions; (iii) informing us about the roles that distinct moral intuitions played evolutionarily and psychologically in human development; and (iv) revealing how our intuitions are subject to the influence of personal characteristics.

Regarding the first way, there is an extensive body of evidence that shows that our moral intuitions are subject to a wide variety of framing effects.[12] As a result, philosophers have started arguing that we should discard moral judgments that we have good reason to suspect are distorted by morally irrelevant factors. They claim that considered judgments should be held on the basis of undistorted, unbiased reasons. Thus, it is useful to learn whether there are conditions under which our judgments about justice are distorted by morally irrelevant factors.

In light of these framing effects, Sinnott-Armstrong (2005) argues that we are becoming more and more capable of distinguishing which intuitions are originated through reliable mechanisms and which are not. Moreover, Sinnott-Armstrong claims that we are now aware of the fact that not all of our intuitions are readily reliable. As a result of this awareness, the author makes a case for the permanent need of confirmation of our intuitions before we can confidently rely on them for the purposes of normative theorizing.

The second manner in which the incoherence amongst folk intuitions can shed light on philosophical issues is by revealing the distinctive psychological and neurological mechanisms responsible for its origin. In this regard, Haidt contends that psychological research has been revealing that it is an emotional process that ultimately generates our moral judgments. Haidt (2001) takes this reasoning even further, maintaining that:

[12] Classic examples of such distortions are illustrated in Kahneman & Tversky, ‘‘Choices, values, and frames.’’ *American Psychologist*, 39 (1984): 341–350.

> Reason can let us infer that a particular action will lead to the death of many innocent people, but unless we *care* about those people, unless we have some *sentiment* that values human life, reason alone cannot advise against taking the action. (p. 345)

In a related vein, Greene (2001, 2004) gathers evidence from neuroimaging that corroborates Haidt's findings. Greene shows that our deontological moral judgments are associated with the activation of brain parts responsible for our emotions. Additionally, he shows that different parts of the brain are activated when we engage in consequentialist moral judgments. Under this latter case, the parts that are activated the most are the ones associated with rational cognition.

The third way in which the incoherence of folk intuitions illuminates our understanding of morality is by informing us about the roles that distinct moral feelings played evolutionarily and psychologically in the course of human natural history. Here again Greene presents us with an interesting line of reasoning based of his empirical findings. He contends that our deontological judgments are in place so as to enable us to live in groups and cooperate with one another. Yet, evolutionarily, these judgments are only fit for small-scale societies–the ones in which we have been living in for the greater part of human history. In his own words, Greene (2008) says:

> I believe that consequentialist and deontological views of philosophy are not so much philosophical inventions as they are philosophical manifestations of two dissociable psychological patters, two different ways of moral thinking, that have been part of the human repertoire for thousands of years. (p. 37)

The fourth and final manner in which the incoherence of folk intuitions can enlighten political philosophical theories is by revealing personal biases in existing approaches. That is, the data can show that allegedly accurate intuitions reached through the traditional armchair process are actually the result of a psychological distortion. Empirical research has shown that individual's conceptions of justice tend to be related with personal characteristics. For instance, Alesina & Giuliano (2009) report that more educated individuals tend to be more averse to redistributive policies, while the opposite holds for women, blacks and respondents with a history of unemployment, or those who were raised Catholic or Jewish. This evidence signals the necessity of making a conscious effort to be aware of these sources of biases, so that philosophers can at the very least try to avoid them. Nagel points in a similar

direction when he argues that individual personal characteristics flavor every great philosopher's version of reality:

> (…) philosophical ideas are acutely sensitive to individual temperament, and to wishes. Where the evidence and arguments are too meager to determine a result, the slack tends to be taken up by other factors. The personal flavor and motivation of each great philosopher's version of reality is unmistakable. (Nagel, 1986, p. 10)

Hence there is no good argument for the dismissal of folk intuitions. The claim that laypersons are not capable of arriving at considered judgments, while expert philosophers would enjoy this capacity, is not defensible. As maintained in this subsection, we hold good reasons for paying due attention to folk intuitions when developing political philosophical theories.

(ii) Against the 'Naturalistic Fallacy Critique'

In this subsection I will address the arguments in favor of the dismissal of the natural fallacy critique. Once again, I emphasize that this group of arguments does not in any manner imply a refutation of the natural fallacy; the logical claim it states remains valid. The arguments that are exposed in this subsection are only intended to refute the use of the natural fallacy as an impediment to empirical and normative collaboration. The arguments that will be respectively examined in this subsection are: *The Feasibility Argument*, *The Public Support Argument*, *The Translation Argument*, *The Measurement Argument*, *The Motivational Argument*, *The New Insights Argument*, and *The Complementation Argument*.

The Feasibility Argument

Schleidgen et al. (2010) argue that Hume's Law logically substantiates the boundaries of empirical-normative collaboration in philosophy, while the Kantian "ought implies can" principle clarifies its particular prospects. They refer in this clarification to the first argument one can make for an active collaboration between the empirical sciences and normative theories: the so-called 'feasibility argument'. Notably, contemporary political philosophers hardly ever deny this argument.

The clarification made by Schleidgen et al. regards the necessity of better understanding human moral behavior so as to demarcate the realm of possibilities for the behavioral dictates of political philosophical theories. In this sense, it is imperative to comprehend what we as humans are *capable of* doing before we establish what we *should* be doing–can must precede ought. As Schleidgen et al. (2010) emphasize, "it is not sufficient for moral norms to demand acts which are logically possible, but empirically impossible due to factual incapacities of moral real subjects" (p. 8).

In this first manner of collaboration, the empirical social and natural sciences can contribute to normative theorizing by helping political philosophers to: (i) specify internal cognitive and motivational capabilities and limits of human agents; (ii) understand externally determined conditions, which are the basic conditions of specific situations which structure the range of possible actions but cannot be influenced by the agents; and (iii) answer questions like how agents actually act in certain situations (which is important in order to evaluate the viability of the norm). Even if we decide to stick with a norm that is initially not viable, it is still important to understand as well as possible how difficult it will be to change human behavior so as to fit the norm and in which ways this change can be achieved.

Hence empirical evidence is crucial at least insofar as political philosophical theories aim at providing guidance for real institutions in real world situations. Once one has a theory of justice and its principles, how can one be sure that people will actually be capable of abiding by them? As nicely stressed by Gaertner & Schokkaert: "Thinking about the content of justice without the desire of making the world more just, is like pouring out a glass of water and then refusing to drink" (2010, p. 8).

The Public Support Argument

The second argument in favor of interdisciplinary research in political philosophy can be called the 'public support argument'. If principles of justice are to serve as guidance for the implementation of public policies, it is of the utmost relevance that these principles share the support of the general population. The fact that this support is directly dependent on the folk's values and preferences makes it essential for a political philosopher to know what these values and preferences are and

to understand as much as possible how they originate and how they evolve over time.[13]

Even if the political philosopher is not going to have his convictions a bit shaken due to the fact that no one shares his considered judgments about justice–already a difficult pill to swallow–it is still paramount to know that such is the case. This relevance is due to the fact that this widespread rejection will be a measure of the likelihood with which policies based on those principles will be effective in the real world.

Alesina and Angeletos (2002, 2004) provide an interesting example of such relevance. Their research focuses on the reciprocal influence between social values and economic policy. The authors show that the values that people hold about social justice matter for policy makers insofar as these values exert direct influence on the levels of government social expenditure. At the same time, the levels of social expenditure implemented by governments also matter for political philosophers insofar as they directly affect the beliefs about justice held by the folk. In this sense, the relevance of empirical evidence is undeniable. As once again nicely framed by Gaertner & Schokkaert, "Even if one considers the majority opinions to be ethically unacceptable, one still has to convince a sufficient number of citizens if one wants to implement one's own supposedly superior conception of justice" (2010, p. 9). In a democratic State, folk intuitions can be shaped and molded, but they cannot be bypassed altogether.

The Translation Argument

The third argument for the use of empirical research in ethics can be called the 'translation argument' (Schleidgen et al., 2010). It states that empirical data should be used as a means to the translation of more general and abstract principles into specific and action-driven directives and guidelines that are both morally justified and workable in practice. The translation argument diverges from the feasibility argument because it claims that empirical data is only relevant *after* the basic principles are already in place; the only parts of the theory that can therefore be questioned by empirical findings are the so-called *practice rules*.

[13] "Empirical research on the acceptance of notions of justice by different social groups is therefore essential to understand the social environment in which policy decisions are taken." (Gaertner & Schokkaert, 2010, p. 8)

Schleidgen et al. (2010) argue that, when dealing with moral justifications for basic principles, it is best to focus on fundamental and systematic analysis, not on empirical issues. They advocate for two levels of analysis: (i) one level that explores the basic principles, which are "pure" and the development of which is the task of normative theorists alone; and (ii) one level that explores the practice rules, which should be empirically informed and tested.

The idea is that normative conclusions have to be translated into practice in accordance with its specific context and conditions. In this sense, the basic principles have to be translated into practice rules so as to come to terms with the specific limits of human thinking and acting. The acknowledgment of this necessity poses a problem to the process of deriving practice rules exclusively from ideal conditions or ideal agents: the real world is not ideal and real people have cognitive and motivational limitations. Hence the derivation of normative practice rules has to be informed both by the ways in which the real world is not ideal and by our knowledge of people's cognitive and motivational limitations. In the following passage, Schleidgen *et al.* (2010) add that:

> However, empirical analysis is neither part of the process of developing a moral norm nor included in the methodological repertoire of normative sciences. Hence, normative theory must rely on collaboration with empirical social sciences (a) when translating basic principles into practice rules and (b) when clarifying the criteria for applying a moral norm. (Schleidgen et al., 2010, p. 5).

Under this view, basic principles should only get 'empirically dirty' when they include the so-called bridging principles. A bridging principle assumes the following form: an action A is demanded in accordance with a moral norm N iff criterion C is met; whereby criterion C must be tested empirically. This means that the conditions of applicability of the principle must be tested empirically.

We can find several examples of bridging principles–which can also be understood in terms of implementation conditions, such as: (1) All sentient beings should not be inflicted pain; the implementation condition is that the being in question is sentient, and this is an empirical claim; and (2) If acting according to N helps in stabilizing society one should act according to N; the implementation condition is that the norm N actually helps stabilizing society, and again this is an empirical claim.

Yet it seems questionable if the "implementation conditions" determine only whether the norm is applicable or not. Sometimes the implementation conditions seem to be determining whether the norm is *actually* valid or not. Some principles only make sense if the world is constituted in some specific way rather than others. For example, taking an extreme case, rational principles only make any sense if we are actually rational creatures. Respecting the rights of individuals only make sense insofar as individuals actually have rights–for example, look at the debates about natural rights and economic rights, or new rights such as the right to labor. In the case of Rawls, for instance, the difference principle is only valid if in fact incentives are needed so as to make people work harder.

The Measurement Argument

The fourth argument can be called the 'measurement argument'. It states that empirical data is significant because it helps us to grasp, describe, and explicate collective processes and changes, which in turn help us to measure the effects of certain norms or rules on the actual performance of agents. This measurement is especially important for the implementation of consequentialist principles insofar as their implementation is dependent on the various effects of distinct alternatives.

The Motivational Argument

The fifth argument can be called the 'motivational argument'. It expresses the importance of psychological knowledge about the nature of human motivation. If justice principles are to have any real effect in the world, they should specify rules such that real individuals are motivated to follow. Sometimes empirical research may reveal why people diverge from moral norms while being at the same time cognitively able to agree with them. In this way, it may be possible to open novel approaches to motivate people to observe these rules (Schleidgen et al., 2010, p. 12).

The New Insights Argument

The sixth argument can be called the 'new insights argument'. It highlights a different manner via which the empirical sciences can contribute to the development of ethical theories: by providing political philosophers with new insights, puzzles, and ideas, which may inform and change their theories. One example of such contribution can be found in the work of Yaari & Bar-Hillel (1984). These researchers provide

evidence for different perceptions of justice about the distribution of goods, depending on whether the distribution is characterized in terms of needs or in terms of tastes. This difference was not accounted for by welfarist theories of justice, and the evidence helped theorists to improve their comprehension of the subject.

Another example of this type of contribution is illustrated by the Pigou-Dalton principle, which states that every transfer of income from a richer to a poorer person that does not reverse the original income ranking of the two individuals is inequality decreasing. Amie & Cowell (1999) showed that a large parcel of the population does not accept this principle. One could have interpreted this as evidence of the folks' stupidity, but Ebert (2009) decided instead to take this evidence seriously. Ebert's work led to the development of the principle of concentration (formerly introduced by Kolm, 1996), being followed by the reinterpretation of the idea of relative deprivation by Magdalou & Moyes (2009).

In addition to providing novel insights in these and related ways, empirical research is also capable of pointing to new facts about the world–such as technological innovations, which demand new or revised principles (Schleidgen et al., 2010). A prominent example of this sort of normative revolution initiated by changes in the world is the emerging field of neuroethics. Before the recent rise of brain research this new field would be unimaginable. Yet today it is one of the most promising areas of investigation in ethics.

The Complementation Argument

The seventh argument can be called the 'complementation argument'. It stresses yet another way in which the empirical sciences can collaborate with normative philosophy, namely, through the complementation of ethical theories. That is, empirical evidence may be needed so as to fill in the gaps of political theories.

There is one paradigmatic example of this sort of collaboration: Roemer's theory of equality of opportunity (1998). Roemer advocates for what is known as 'leveling of the playing field', arguing that every person is entitled to an equal chance to succeed. In order to achieve this equality while at the same time making it possible for persons to reach different levels of success, he builds on the classical distinction between effort and circumstances. Individuals are to be held responsible for the former, and compensated for the latter.

There exists a world of philosophical debate as to where the line between luck (social and natural lottery) and effort (taken to be within the realm of personal control) should be drawn (Fleurbaey, 2008), but Roemer has a different view on the subject. He states that the line is culture-dependent: "Because the choice by society of these parameters cannot but be influenced by the physiological, psychological, and social theories of man that it has, the present proposal would implement different degrees of opportunity egalitarianism in different societies" (Roemer, 1993, p. 166). Hence he leaves an open invitation for empirical work on cultural disparities in the levels of responsibility attribution. His theory offers a general and coherent framework that can be applied for any division between effort and circumstances, while empirical work supplies the necessary information about where the boundary is to be drawn in different societies.

Another example by Gaertner & Schokkaert addresses what should be done when we face a conflict of valued interests between, for instance, generating economic growth and violating individual rights (such as the right to strike). They claim that:

> A priori ('objective') theories of well-being (such as the one by Nussbaum, 2000, 2006) might offer a framework for dealing with the resulting trade-offs, but even these theories often remain silent about the structure of relative weights and are therefore not very helpful in specific situations. Another approach, which is much more in line with the economic tradition, is to respect ('subjective') individual opinions and preferences about these trade-offs. In this latter approach, empirical work is needed to collect the necessary information about preferences. (Gaertner & Schokkaert, 2010, p. 13)

In this case, it is important to stress that the principle of respect for preferences is *a priori* and therefore needs philosophical justification. Nonetheless this necessity does not eradicate the need of empirical work to provide this principle with practical substance. In this sense, Gaertner & Schokkaert (2010) provide a series of cases in which the general public has clear preferences about trade-offs in relation to which theories have not provided clear guidance. They claim that in these cases empirical evidence can provisionally provide this necessary guidance, at least until we have a complete normative theory.

4. Reflective Equilibrium and Public Justifiability

There is yet a more substantial–and ambitious–claim in favor of taking empirical work seriously in the development of political theories. This claim is related to a particular way of doing political philosophy, one that is exemplified by the valuable Rawlsian conceptions of reflective equilibrium and public justifiability.

Firstly, it is worth stressing once again that moral intuitions play a crucial role in Rawlsian reasoning. As John Rawls writes, "One may regard a theory of justice as describing our *sense of justice*. (…) A conception of justice characterizes our moral sensibility when the everyday judgments we do make are in accordance with its principles."[14] That is, principles of justice must emerge from a balance between some of our principles and some of our intuitions and considered judgments.

This is not to say that a theory of justice is merely a catalogue of folk intuitions. Here is where the conception of reflective equilibrium enters into the scene. Moral intuitions are important insofar as they are the starting point to the process of achievement of a narrow reflective equilibrium. As nicely delineated in the Stanford Encyclopedia of Philosophy:

> In carrying through this method one begins with one's considered moral judgments: those made consistently and without hesitation when one is under good conditions for thinking (e.g., "slavery is wrong," "all citizens are political equals"). One treats these considered judgments as provisional fixed points, and then starts the process of bringing one's beliefs into relations of mutual support and explanation as described above. Doing this inevitably brings out conflicts where, for example, a specific judgment clashes with a more general conviction, or where an abstract principle cannot accommodate a particular kind of case. One proceeds by revising these beliefs as necessary, striving always to increase the coherence of the whole. Carrying through this process of mutual adjustment brings one closer to *narrow reflective equilibrium*: coherence among one's initial beliefs.

After we have achieved this narrow reflective equilibrium, we proceed to the process of wide reflective equilibrium. We engage in this second stage by adding to our responses the major theories in the history of political philosophy, as well as the theories that are critical of political philosophizing as such. We continue to make

[14] Rawls, A Theory of Justice, 1971, p. 41.

adjustments in our schemes of beliefs as we reflect on these alternatives, aiming for the end-point of *wide reflective equilibrium* in which coherence is realized after many alternatives have been considered.

As Rawls emphasizes, the best account of a person's sense of justice is one that "matches his judgments in reflective equilibrium."[15] The idea is not to simply read off principles of justice from common sense moral judgments–but these judgments nevertheless serve as important inputs into the process. Moral intuitions must be filtered by a procedure of impartial reflection. That is, we seek an account that systematizes, in Rawls's terms, our considered moral judgments.

Moreover, a person may be right to accept a theory of justice that fails to accommodate some of her considered moral judgments. She may decide that this theory does an otherwise admirable job of explaining her most highly esteemed considered judgments. Hence she chooses to revise or discard the particular considered moral judgment that conflicts with the theory rather than to revise or discard the theory.

We bring considered moral judgments into reflective equilibrium by undergoing a process of revising general principles against particular judgments. We discard a general principle if it yields a particular judgment we refuse to accept; we discard a particular judgment if it violates a general principle we refuse to revise. Eventually we reach a satisfactory balance of principles and judgments. Thus, the principles of justice are not meant to serve as *ad hoc* explanations of our common sense intuitions. Our goal is to arrive at a systematic articulation of the verdicts of moral *common sense*. These principles bring out the so-called *deep structure* of our moral beliefs (Miller, 2003).

Miller stresses the importance of a second Rawlsian core idea, contending that the possibility of justifying a theory of justice to the general public is a precondition for its being ethically acceptable–the so-called public justifiability argument. The goal is to ensure that all valid principles of justice will be capable of being publicly justifiable. That is, valid principles must be such that citizens of a well-ordered society can justify them to one another using only commonly accepted modes of argument.

In this case, Miller insists that:

---

[15] Ibid, p.43.

> It seems much more plausible to regard the set of beliefs that are publicly justifiable in a given society S as the beliefs currently held in S adjusted to take account of empirical error, faulty inferences, the distorting effect of self-interest, and so on–that is, the deficiencies that are already commonly understood to produce erroneous beliefs.
> (2003, p. 56)

Gaertner & Schokkaert agree with this view, clarifying that:

> Views such as the one of Miller certainly do not conflate social scientific research on justice with normative theory. Popular vote is not the ultimate justification of an ethical position. Opinions of the public are no more than an input (albeit a necessary one) into a broader philosophical debate aiming at a reflective equilibrium between theoretical principles and specific considered judgments. Putnam gives a larger weight to majority opinions, but also in his view there remains an essential tension between public opinion and normative thinking. (…) Therefore, the role of theoretical thinking remains essential. Yet, in these approaches, theoretical thinking should necessarily integrate in a critical way the findings of empirical work. The latter therefore is an essential ingredient into the normative debate. (2010, p. 17)

Rawls continuously remarks that principles of justice should express the fundamental ideas implicit in the public political culture of a democratic society. At the same time, he repeatedly states that when a principle is tested via reflective equilibrium the only opinions that count are those of the philosopher and of the reader of his book. Hence Rawls is pulled in different directions when it comes to empirical evidence: he simultaneously adheres to a form of contractarian reasoning (which does not rely on empirical evidence) while relying on judgments that are supposedly shared by the general public.

Thus we are left with the following question: is it possible to decide whether a judgment is considered simply by scrutinizing it in solipsistic fashion, relying only on internal evidence to establish how much confidence we should place in it, or whether it has been influenced by one of the distorting factors that Rawls mentions? It is surely of the greatest relevance to check whether the judgments we make are shared by those around us, and if they are not, to try to discover what lies behind the disagreement (Miller, 2003, p. 55).

In this sense, experimental evidence should function as actual guidance to normative theories. That is, we should make use of folk intuitions and beliefs as an

active source of information in order to better understand the content of the principles of justice. This is not to say that we can simply derive normative principles from descriptive ones. It is instead a claim about the nature of ethical beliefs and its objectivity. This argument is nicely developed by Amartya Sen in his work *The Idea of Justice*:

> (…) public reasoning is clearly an essential feature of objectivity in political and ethical beliefs. (…) In seeking resolution by public reasoning, there is clearly a strong case for not leaving out the perspectives and reasonings presented by anyone whose assessments are relevant, either because their interests are involved, or because their ways of thinking about these issues throw light on particular judgments – a light that might be missed in the absence of giving those perspectives an opportunity to be aired.
> (2009, p. 44)

If theories of justice are to articulate our shared conception of justice–in Rawls's terms, a conception "which is congenial to the most deep-seated convictions and traditions of a modern democratic state"–we should conduct empirical research to learn which conception of justice is actually shared by the citizens of modern liberal democratic states.[16] We cannot simply assume from the armchair that the philosophers' intuitions are representative of the intuitions of laypersons. Claims about the distribution of intuitions are ultimately empirical claims. Thus, Miller highlights that "in setting out a theory of justice, the normative theorist who is guided by something akin to the Rawlsian ideas of reflective equilibrium and public justifiability needs evidence about what people do in fact regard as fair and unfair in different social settings," reckoning that "a theory of justice needs to be grounded in evidence about how ordinary people understand distributive justice."[17]

5. Final Considerations

We are now living under a new paradigm whereby we are beginning to better understand how our brains operate.[18] As a result, we are becoming increasingly capable of developing political philosophical theories for *real* institutions and

[16] John Rawls, *Political Liberalism*. New York: Columbia University Press, 2005, p. 300.
[17] *Principles of Social Justice*, pp. 59; 61.
[18] The novel findings from behavioral and brain research will be properly addressed in the second paper that composes this dissertation.

persons. Moreover, as we have discussed, the empirical sciences provide an array of relevant data about human beliefs and behavior that can inform political philosophers in a variety of ways.

In this context, Weaver and Trevino (1994) envision three possible ways in which science and normative philosophy can actively collaborate. The first way is the so-called *symbiotic collaboration*. The symbiotic type of interdisciplinarity advocates for a relation amongst ethics and the empirical sciences in which one supplements the other in its limitations. That is, the symbiotic approach entails a pragmatic and collaborative relation between normative theorizing and empirical research, in which the cores of each approach remain essentially separated.

The second approach for the collaboration of empirical sciences and normative theories is the so-called *parallel*. This approach implies the utter denial of any possible integration between empirical and normative theories, on both conceptual and practical grounds. Advocates of this approach argue for the strict separation between that which is normative and that which is descriptive. As they emphasize, normative theorists and empirical scientists should work as 'parallel lines'; they should never allow their researches to 'touch' each other.

The third and final manner of collaboration is the so-called *integrative* approach–which rejects the very idea of a distinction between normative and empirical claims. The supporters of the integrative approach go even further, stating that 'there is no fundamental distinction between fact and value' or 'between descriptive and prescriptive science' (Molewijk et al., 2004). Under this approach, it is understood that empirical research about normative practices are able to generate normative philosophical theories (van der Scheer & Widdershoven, 2004).

As I have argued throughout this paper, the parallel approach lacks significant support. Therefore, we are left with the remaining two forms of collaboration: the symbiotic and the integrative approaches. The latter constitutes a bold claim that may not sound as implausible as one might think; yet, it would require more arguments in its favor than I was able to consider in the preceding sections. Hence, we are left with the former, a fruitful approach that fits well with the ideas developed in this work.

According to the symbiotic approach, political philosophers can no longer afford to ignore all the relevant empirical information from the natural and the social sciences in the development of their theories. Normative theories ultimately concern the *actual* behavior of *real* institutions and *real* human beings in the *real* world, not

the *assumed* behavior of *idealized* institutions and *idealized* human beings in *hypothetical* worlds.

Theories of Distributive Justice and Experimental Evidence: An Interdisciplinary Review of Contemporary Data on the Concept of Justice

> Rawls does not conceive of moral philosophy as depending primarily on the analysis of valid moral argument. Rather, he thinks of a theory of justice as analogous to a theory in empirical science. It has to square with what he calls 'facts', just like, for example, physiological theories. *But what are the facts?*
>
> Hare (1973, p. 145)

Introduction

Political philosophers are broadly concerned with the study of human social organization.[19] More specifically, they aim at the elaboration of a set of principles capable of outlining how we should organize our social institutions so as to be able to live not as atomistic individual beings, but as active members of cooperative endeavors.[20] How should we understand our mutual responsibilities to one another as members of a society? What sorts of treatments do we rightly owe each other?

In the search for the principles that will provide the answers to the above and related questions, the main virtue in which political philosophers are interested is justice–the primary virtue of social institutions (Rawls, 1971). And within the realm of justice emerges one of the central areas of research in political philosophy and the focus of this dissertation, namely, distributive justice. Principles of distributive justice are meant to guide the workings of the main social institutions with respect to the allocation of burdens and advantages among the members of a society, such as the allocation of education, medical treatment, and taxes. A theory of distributive justice is crucial to the development of a fair and well-functioning society, even if this theory

[19] In the *Companion to Contemporary Political Philosophy* one finds the following definition: "(…) what is the concern of political philosophy? Primarily, it is a concern to identify the sorts of political institutions that we should have, at least given the background sort of culture or society that we enjoy. To take the view that we should have certain political institutions will imply that if such institutions are in place, then, other things being equal, agents should not act so as to undermine them" (2012, pp. xvi-xvii).

[20] It is worth noting that this enterprise does not stop some philosophers from envisioning these principles as entailing an atomistic view of society.

is tacit and not explicitly developed, as in small ancient societies. Moreover, in order for a theory of distributive justice to be successful it must be able to "persuade people to regulate their intuitive sense of justice by its principles and allow this hope to be realized" (Miller, 2003, p.21).

In spite of the central role played by theories of distributive justice in contemporary political philosophy,[21] there is to date nothing close to a consensus on which principles should guide the allocation of social and economic benefits and burdens amongst individuals. All the more disturbing, as nicely highlighted by Miller (2003):

> (…) contemporary liberal moral and political philosophy presents a spectacle of continuing deep disagreement between rival theories of justice. Each theory claims to embody demonstrable truth, but there is no reason to think that the contest between them will ever be resolved. (p.112)

In the face of the practical relevance of the subject and the present state of 'comprehensive dissent' about which sort of distributive justice principles we should adopt, it is important to address the question: why do we find ourselves in this current state of "deep disagreement"? There is no doubt that the answer is partly related to the complexity of the matter–justice is indeed not straightforward! Yet attributing the problem solely to the complexity of the subject would be an acknowledgment of inevitable failure, an acceptance of the impossibility of the pursuit. Therein rests the necessity of exploring other reasons.

A promising road for the undertaking of this task is to investigate one common aspect of contemporary theories of distributive justice. This aspect is one that goes back to Descartes and has a short discontinuity during the English Empirical Enlightenment, namely, contemporary political philosophers' methodological approach. The usual method of so-called *armchair philosophy* is the method of abstraction from intuitions: the idea is to start from what philosophers *claim* to constitute people's basic intuitions and from that build a rationally coherent set of

[21] "(…) what is connoted by our focus on *contemporary* political philosophy? Within the analytical tradition of thought, as that affects both philosophy and other disciplines, political philosophy has become an active and central area of research in the past three or four decades; it had enjoyed a similar status in the nineteenth century but had slipped to the margins for much of the twentieth. In directing the *Companion* to contemporary political philosophy, we mean to focus on this recent work" (*Companion*, 2012, p. xvii).

abstract principles. In contemporary political philosophy, this method has been altered and perfected by Rawls in his conception of *reflective equilibrium*.

For Rawls, the most abstract aim of political philosophy is to reach justified conclusions about how political institutions should be arranged. He understands the degree to which one is justified in one's political convictions as dependent on how close one is to achieving *reflective equilibrium*. In reflective equilibrium all of one's beliefs, on all levels of generality, cohere perfectly with one another. Thus, in reflective equilibrium, one's specific political judgments support one's more general political convictions, which in turn support one's very abstract beliefs about oneself and one's world. Viewed from the opposite direction, in reflective equilibrium one's abstract beliefs explain one's more general convictions, which in turn explain one's specific judgments. Were one to attain reflective equilibrium, the justification of each belief would follow from all beliefs relating in these networks of mutual support and explanation.

Though perfect reflective equilibrium is unattainable, we can use the *method of reflective equilibrium* to get closer to it and so increase the justifiability of our beliefs. In carrying through this method one begins with one's considered moral judgments: those made consistently and without hesitation when one is under good conditions for thinking. One treats these considered judgments as provisional fixed points, and then starts the process of bringing these beliefs into relations of mutual support and explanation as described above. Doing this inevitably brings out conflicts where, for example, a specific judgment clashes with a more general conviction, or where an abstract principle cannot accommodate a particular kind of case. One proceeds by revising these beliefs and principles as necessary, striving always to increase the coherence of the whole.

Carrying through this process of mutual adjustment brings one closer to *narrow reflective equilibrium*: coherence among one's initial beliefs. One then adds to this narrow equilibrium one's responses to the major theories in the history of political philosophy, as well as one's responses to theories critical of political philosophizing as such. One continues to make adjustments in one's scheme of beliefs as one reflects on these alternatives, aiming for the end-point of *wide reflective equilibrium* in which coherence is realized after many alternatives have been considered. Current liberal egalitarian philosophers seem to have sided with Rawls when it comes to the proper manner of developing their theories.

Following the stream of Rawlsian ideal theorizing, contemporary political philosophers of the liberal tradition have been reluctant to take several sorts of empirical evidence seriously in the development of their theories of distributive justice. This claim does not amount to stating that political philosophers have ignored *all* kinds of empirical facts. Quite the contrary, as will be discussed onwards, they have been rather attentive to a large body of evidence from the applied social sciences. Rawls, in particular, has always been acutely sensitive to the findings from the social sciences. Nonetheless contemporary political philosophers have not to date accounted for a large portion of relevant empirical results, particularly those from the natural sciences.

There are several reasons[22] for this cautious attitude towards the empirical, yet two of them seem most salient: (i) the fear of falling into a descriptive account of morality, with no normative power; and (ii) the allegation that folk intuitions are unable to adequately reveal an individual's *considered judgment*s. Nonetheless, as I have argued in a separate paper, neither of these worries is justified.[23] Indeed, on my view, there are an over-determining number of reasons[24] for adopting precisely the opposite attitude, i.e., for embracing the incorporation of all *relevant* empirical evidence into political philosophical theorizing about distributive justice. Just to mention one sufficient reason for adopting this attitude, consider the lack of practical guidance provided by distributive principles that are dissociated from real patterns of human behavior and beliefs.

In view of the purpose of principles of distributive justice, namely, determining the fair allocation of valuable resources amongst persons, what type of empirical evidence would be of relevance for political philosophers? A seemingly appropriate place to start looking for this answer is the *Companion to Contemporary Political Philosophy*–a work that intends to provide in-depth coverage of the contemporary political philosophical debate. As stressed by the editors,

> We would also like to think that, without any heavy-handed attempt on our part at imposing uniformity on what is by its nature a disparate academic community, our contributors have managed among themselves to produce a genuinely coherent synopsis of the 'state of play' in contemporary political philosophy worldwide. (2012, p. ix)

[22] These reasons (and the arguments for and against them) have be been addressed in the first paper.
[23] Vide first paper of this dissertation.
[24] Vide paper abovementioned.

In relation to the relevance of empirical evidence, one important feature of the *Companion* is the emphasis on a broad view about the range of issues that are normatively relevant to political philosophy. The editors assume that "questions about what can feasibly be achieved in a certain area are just as central to normative concerns as questions about what is desirable in that area," going on to add that they "understand political philosophy in such a way that it does not belong to the narrow coterie of those who would just contemplate or analyze the values they treasure" (2012, p. xvi).

In relation to the kind of empirical evidence that is of relevance to political philosophers, one important feature of the *Companion* is the coverage of a range of disciplines that have contributed to the development of political philosophical views. In this respect, the editors "look, not just to philosophy–analytical and continental–but also to economics, history, law, political science, and sociology" (2012, p. xvi). Yet, the focus of the *Companion*, as narrow as it is in covering only contemporary political philosophy, is still broader than my focus in this work. I do not intend to explore empirical evidence relevant for theories of democracy, voting behavior, etc. Here my center of attention is exclusively on theories of distributive justice and, as a consequence, I will be interested in the empirical data relevant for the development of these theories.

It is important to stress that the fact that in the *Companion* we encounter an array of applied disciplines is in itself quite revealing. The presence of disciplines such as economics and social biology reveals that an empirical turn has already started in the field. In this context, my goal is to broaden the scope of this empirical awareness.

Hence, in order to ascertain which kinds of empirical evidence are of relevance for political philosophers interested in theories of distributive justice, I will look into an array of different areas of research that have to date not been properly acknowledged by political philosophers. These fields include distinct empirical sciences that have been developing research related to human behavior and the concept of justice such as neuroscience, evolutionary biology, social psychology, and experimental economics.

It is imperative to clarify that political philosophers cannot be so simply accused of ignoring these new empirical developments. The first reason for not being

so fast in blaming philosophers for denying the relevance of these relatively new sciences is the mere fact that they are *new*. Therefore, it is to be expected that some time will be required before political philosophers can properly incorporate these new findings.

Nonetheless, this novelty does not completely absolve contemporary political philosophers. After all, moral philosophers have already been attentive to these empirical fields. What is more, they have themselves joined the effort of providing specific relevant empirical evidence for their theories. For instance, moral philosophers, in collaboration with moral psychologists, have in the past two decades gathered significant data indicating that our moral rules are more emotional than the rationalist crowd would have expected. The growing literature suggests that our moral judgments are triggered by emotional reactions, and that we are easily *morally dumbfounded* by our own moral intuitions (Haidt et al., 1993). Numerous other experiments have shown that our moral judgments are strongly affected by environmental cues, heuristics and biases, emotional intuitions, and the like (e.g. Cushman et al., 2006; Greene et al. 2004; Sinnott-Armstrong et al., 2010; Wheatley & Haidt, 2005). And this is just to mention one stream of empirical data that has been recently emerging in the literature.

At this point, another clarification becomes imperative. The embracement of the relevance of empirical methods in political philosophy entails neither: (i) the rejection of traditional philosophical methods; nor (ii) the necessity of having political philosophers doing empirical work themselves. Instead the proper acknowledgment of the role of empirical data in political philosophy merely involves: (i) the need for political philosophers to always avail themselves of the results of the empirical sciences that are relevant to their focus of interest; and (ii) the collaboration of philosophers and other scientists in empirical research whenever necessary and academically profitable for the development of their respective fields.

In this vein, the aim of the present paper is to provide an extensive review of the existing empirical literature on human intuitions, beliefs, and behaviors related to the concepts of justice and fairness. This review includes some of the most significant research involving these concepts during the past three decades in the areas of primatology, evolutionary biology, experimental economics, moral psychology, political and social psychology, and neuroscience. The idea, once again, is to advocate for the value of interdisciplinary work in political philosophy, an area that is

by nature multidisciplinary – and should therefore be treated as such. Additionally, I hope that making all these novel research programs and some of their interesting results easily available for political philosophers will fuel the development of an empirically informed political philosophy.

The paper is structured in three sections. The second section following this introduction offers a presentation of the relevant empirical data from the fields of primatology, evolutionary biology, experimental economics, moral psychology, social and political psychology, and neuroscience. The third section, to conclude, provides a discussion of some of the implications of these findings for political philosophy and, more specifically, to theories of distributive justice, considering possible roads for future research.

2. The Empirical Evidence

In this section I will present an extensive review of the empirical literature on the conception of justice. Some of the research fields covered by this review have already been established by the literature, others are emerging areas of investigation that have been revealing novel results and illuminating data on human morality.

In the preceding paper I claim to have established the relevance of the results from the empirical sciences for proper political philosophical theorizing about distributive justice. Now, as stressed in the introduction, my aim is to make these empirical findings available for philosophers so as to fuel better informed justice theories. The experimental findings reported here are organized in the following subsections: (i) Findings from Primatology; (ii) Findings from Evolutionary Biology; (iii) Findings from Experimental Economics; (iv) Findings from Moral Psychology; (v) Findings from Social and Political Psychology; and (vi) Findings from Neuroscience.

(i) Findings from Primatology

Principles of justice, whether implicit or explicit, are interpreted by numerous scholars of different areas as a relevant feature for the stability of social systems. As stressed by Garret Hardin, "the first goal of justice is to create a *modus vivendi* so that life can go on, not only in the next five minutes, but also indefinitely into the future"

(1983, p. 412). It is widely known that humans are not unique in organizing life as a joint endeavor; several animal species organize their lives within the boundaries of societies. As communities, these animals share rules of conduct in order to make coexistence both possible and successful; i.e., so as to achieve the so-called stability of their social systems. And this is only one of the many features shared between humans and other species, one that points to the enlightening possibility of the existence of proto-moral systems beyond the narrow scope of human societies.

In order to shed light on this issue, what better species with which to begin this investigation than great apes and monkeys? This is precisely what primatologists like Frans de Waal have been doing. They have been studying behaviors analogous to those observed in humans in other primates. In what concerns the realm of ethical behavior, their goal is to gather evidence in support of the hypothesis that human morality is (at least partially) a product of natural selection. As emphasized by Fleck & de Waal (2002) "morality indeed may be an invention of sorts, but one that in all likelihood arose during the course of evolution and was only refined in its expression and content by various cultures" (p. 2).

They are fully aware of the controversial nature of their research, yet their claim is not especially lavish. Fleck & de Waal (2002) do not consider chimpanzees to be actual moral creatures, but only to portray what they call elements of moral systems in their societies. Although we may not be able to establish any relations amongst the specifics of our moral systems and biology, the investigation of the degree to which biology has had an effect in the shaping of human moral systems remains a worthwhile pursuit.

For instance, primatologists have collected data that indicates the presence of similar methods, amongst humans and other primates, for managing conflicts of interests within groups. These methods include practices such as reciprocity, food sharing, reconciliation, consolation, conflict intervention, and mediation. Many of these practices constitute advanced methods of resource distribution, which in humans are taken to imply the existence of underlying psychological mechanisms like the capacity for empathy and sympathy – usually considered necessary attributes for moral reasoning. The fact that other primates exhibit similar practices points in the direction of shared psychological mechanisms and, perhaps, a shared proto-morality.

Fleck & de Waal (2002) report extensive evidence on the habits of food sharing amongst chimpanzees, bonobos, siamangs, orangutans, and capuchin

monkeys. This behavior is explained by the authors via the "reciprocity hypothesis," that conceives food sharing as an element of a broader system of mutual obligations, which include either the exchange of material goods or of social favors. The novelty of this hypothesis relies on its emphasis in the joint nature of the relation between, for instance, possessors and beggars of food. To follow the example, under this view the sharing of food is regarded as mutually beneficial.

De Waal (1997a) also provides evidence for the existence of calculated reciprocity, arguing that "the possibility that chimpanzees withhold favors from ungenerous individuals during future interactions, and are less resistant to the approaches of individuals who previously groomed them suggests they have expectations about how they themselves and others should behave in certain contexts" (Fleck & de Waal, 2002, p. 8). This evidence is present both in beneficial conflict interventions (when A interferes in a conflict in benefit of B; and B returns the favor, in the future, in likely manner) and in harmful interventions (when C interferes in a conflict against C; and C interferes against A, in the future) (de Waal, 1997a).

Fleck & de Waal (2002) interpret this evidence concerning reciprocity habits as telling us something about the origins of our own sense of justice. As they suggest, these rituals of reciprocity "exemplify how and why prescriptive rules, rules that are generated when members of a group learn to recognize the contingencies between their own behavior and the behavior of others, are formed. The existence of such rules and, more significantly, of a set of expectations, essentially reflects a sense of social regularity, and may be a precursor to the human sense of justice" (Fleck & de Waal, 2002, p. 9).

Another ritual that we share with most primates is reconciliation, observed as a post-conflict behavior, in the form of kindly reunion between previous opponents short after the occurrence of a confrontation. This ritual seems to constitute a universal method of repairing disturbed relationships. About its frequency, it is interestingly observed that "individuals in despotic species reconcile less frequently after conflicts than individuals in more tolerant and egalitarian species, most likely because the strict dominance hierarchies that are present in despotic species constrain the development of strong symmetrical relationships among group members"(Fleck & de Waal, 2002, p. 12).

Regarding sentiments of empathy and sympathy, primatologists have observed the pervasive existence of succourant behavior amid great apes – which includes

behaviors such as taking care and relieving the distress of non-kin individuals. This finding suggests that great apes possess the ability both to be concerned about others and to understand their needs and emotions (Scott, 1971). In this vein, Hatfield, Cacioppo & Rapson (1993) provide evidence of the phenomenon of emotional contagion in infant primates; meaning they have the capacity to become distressed as a consequence of perceiving distress in other individuals. They report that infant primates reveal a necessity to comfort themselves when faced with a fight, despite the fact that they had no involvement in the fight. Additionally, there exists systematic data showing that great apes have the capacity to engage in active consolation. For example, observations have shown that it is not unusual for a juvenile chimpanzee to approach and embrace an adult male who has just been defeated in a fight against its rival (de Waal, 1982).

Based on the present results from primatology, Fleck & de Waal (2002) conceive a categorization of the evolutionary building blocks of human moral systems. It is important to emphasize that all primates share these same building blocks. These building blocks are not constituted by clusters of behaviors characterized by humans as good and nice, but by mental and social capacities that enable societies to hold shared values that constrain individual behaviors by the implementation of a system of approval and disapproval. The categories are: (i) *Sympathy Related* – attachment, succourance, emotional contagion, learned adjustment to and special treatment of the disabled and injured, and ability to trade places mentally with others; (ii) *Norm Related* – prescriptive social rules, internalization of rules and anticipation of punishment, a sense of social regularity and expectation about how one ought to be treated; (iii) *Reciprocity* – giving, trading, revenge, and moralistic aggression against violators of reciprocity rules; and (iv) *Getting Along* – peacemaking, avoidance of conflict, community concern, and accommodation of conflicting interests through negotiation. The Norm Related block is of particular relevance to our topic of interest, especially the sense of social regularity and expectation about how one ought to be treated. Fleck & De Waal interpret both this sense and this expectation as incipient traits of our sense of justice.

In the face of these findings, Arnhart (1998) argues that our biologically based desires and cognitive capacities are responsible for the emergence of our moral systems. According to Arnhart, our moral rules seem to be the product of our sophisticated cognitive abilities, enhanced by the capacity for language, and grounded

on an evolutionarily developed set of natural capacities. Our natural capacities would be realized through social learning and moral habituation, and our specific moral systems would vary in accordance with the specificities of our societies' social and physical circumstances.

All these findings from primatology suggest a naturalistic perspective on human morality, one that can be potentially threatening to the normative status of our moral rules. Moreover, this perspective is largely understood in connection with our evolutionary history – as will emerge more clearly in the next subsection.

(ii) Findings from Evolutionary Biology

Evolutionary theory is concerned with the study of the biological origins of our physical features and, more recently, of our psychological ones. In relation to our topic of interest, i.e., distributive justice, evolutionary theory has been recently providing an astonishing amount of new findings on the evolution of morality and moral systems. The main goal of this research is to provide an answer to the question: is our moral system a product of evolution?

In order to answer this question, researchers have to identify the adaptive functions that our moral system has supposedly evolved to serve. In this sense, it has been argued that the chief function performed by our sense of justice is to provide an incentive for individuals to engage in cooperative activities. Therefore, evolutionary theorists claim that we must understand the emergence of cooperative mechanisms if we want to understand the origins of our sense of justice.

In this vein, one of the major streams of contemporary research in evolutionary theory is concerned with the investigation of the origins of altruistic behavior. Altruistic behavior is defined, in the evolutionary sense,[25] as all behavior that simultaneously involves a cost to the donor and a benefit to the recipient. Both costs and benefits are interpreted in terms of the currency of reproductive success – namely, fitness.

Altruism constitutes a problem from the standpoint of evolution given that incurring fitness costs is obviously counterproductive for the reproduction of genes. One of the principal hypotheses developed to account for this problem is the

[25] It is important to stress that evolutionary biology does not aim at explaining the origins of psychological altruism, which is related to a non-instrumental concern for the welfare of others.

hypothesis of group selection. It explains altruistic behavior as the result of a process of selection for groups. This explanation stands in opposition to the view that natural selection always works either at the level of the individual or at the level of the gene itself. The idea is that groups of individuals compete for survival. In this competition, the presence of altruistic individuals works as an advantage for the group, given that this presence increases the group's likelihood of survival. As a consequence of this beneficial feature, altruism ends up being selected for via evolution. Contrarily, some evolutionary theorists believe only in selection at the individual level, arguing against the hypothesis of group selection. Along this line, George Williams (1966) states that evolution theory explains solely the evolution of selfish traits that evolve due to the promotion of the replication of genes.

Despite the fact that evolutionary theory does not (at least primarily) account for psychological altruism, evolutionary explanations for the phenomenon still remain an open possibility. Sober & Wilson highlight that "psychological motives are proximate mechanisms in the sense of the term used in evolutionary biology. (…) if certain forms of helping behavior in human beings are evolutionary adaptations, then the motives that cause those behaviors in individual human beings also must have evolved" (2002, p. 200)[26]. Sober & Wilson (2002) embrace this line of research and argue for the evolution of a plurality of motivational forces that drives our altruistic behavior. These motivational forces include hedonistic motivation and also a regard for other's welfare. According to the authors, we evolved so as to presently exhibit mixed motivations for our other regarding behavior.

Another important source of evolutionary research into the origins of morality involves the use of game theoretical approaches. Amongst its prominent scholars is Brian Skyrms, who emphasizes the role that game theoretical approaches can play in the understanding of our moral systems. In his words:

> Hobbes wanted to bring the rigor and certainty of Euclidean geometry to social philosophy. If he fell somewhat short of this goal, even by the standards of his own time, perhaps the theory of games could be utilized to complete the project. This idea has been pursued in different ways by John Harsanyi, John Rawls, David Gauthier and Ken Binmore.
> (Skyrms, 2002, p. 269)

[26] There are three principles in biology that support evolutionary accounts of psychological altruism. These principles are: (i) availability; (ii) reliability; and (iii) efficiency.

Game theoretical models of natural selection are "potentially useful to psychologists interested in understanding the acquisition of morality because they may help them decide whether the dispositions to adopt various cooperative strategies could have evolved, and cooperation lies at the heart of morality" (Krebs, 2002, p. 314). In this same direction, biologists have provided significant evidence that norms of reciprocity have evolved, given its presence in non-human animals societies. A prominent example of one such norm is the Tit for Tat strategy. That is, 'an eye for an eye; a tooth for a tooth.' Another example is the Golden Rule: 'do unto others as you would have them do unto you.' This second rule represents an unconditionally cooperative strategy, in contrast with the former, which represents a conditionally cooperative strategy. Krebs (2002) argues that the Golden Rule could have evolved in societies where strategies like Tit for Tat lead to the extinction of selfish strategies. Yet, he also reminds us that "the larger the number of unconditional cooperators, the more fruitful the environment for the invasion of selfish individualism, suggesting that the natural state of moral dispositions in the human species may be unstable and fluctuating" (Krebs, 2002, p. 317).

Tit for Tat and the Golden Rule are strategies that involve direct reciprocity. What about strategies of indirect reciprocity? Could they also have evolved? Richard Alexander (1987) first elaborated the view that our moral systems are more like systems of indirect reciprocity, in which one individual benefits another without receiving anything in return from that specific individual, but ends up being the recipient of some other benefit provided by a third individual in the future. He argues that the Golden Rule could have evolved via the implementation of mechanisms of indirect reciprocity supported by discrimination in favor of altruists and against cheaters.

Alexander's (1987) argument relies on the idea that beneficence pays off in three different ways. Firstly, beneficent individuals could end up being the ones selected as exchange partners within societies. Secondly, beneficent individuals could end up being rewarded by other group members and relatives via status elevation or resource giving. And thirdly, beneficent acts could increase the success of the entire group, therefore also increasing the beneficent individual's share of benefits.

In support of Alexander's (1987) hypothesis, Nowak & Sigmund (1998) show that mechanisms of indirect reciprocity are fully capable of giving rise to different

forms of altruism – including direct reciprocity. They demonstrate this possibility via computer simulations.

It is important to stress that systems of direct and indirect reciprocity are by no means mutually exclusive. This research provides illuminating insights into the role that impartiality plays in the generation of our moral judgments. Krebs (2002) affirms that

> (…) we are biologically disposed to maximize others tendency to practice the Golden Rule in their interactions with us, which we do by preaching this principle, [and] creating the impression we practice it, at least more than we actually do. (…) the stronger our beliefs in our moral worthiness, the better our ability to convince others of our morality, and, therefore, the better we are treated by others. Such self-deception is adaptive. (Krebs, 2002, p. 319)

This self-deception about our moral worthiness is not only adaptive, but it also helps to shape our behavior in a way that ends up being self-fulfilling: the fact that we repeatedly behave in a moral manner to some extent makes us moral. This is the upside of our self-deception. As Krebs points out:

> To sustain mutually beneficial cooperative relations, cultivate a favorable social image, avoid ostracism, and sustain our impressions of ourselves, we must help those who help us and behave altruistically some of the time, at least in public. Although few if any of us are biologically disposed to adopt the unconditional strategy of doing unto others as we would have them do unto us, we may end up behaving in ways that are consistent with this principle with some people some of the time. (Krebs, 2002, p. 319)

Based on this research, Krebs (2002) states that human morality evolved in such a way that we inherited flexible programs that organize sets of conditional strategies. He argues for the existence of strategies that are domain-specific. This domain-specificity means that these flexible programs regulate distinct types of social relations, namely, hierarchical, egalitarian, and intimate. These are interesting findings in light of an increasing number of philosophers currently arguing for pluralist theories of justice. For instance, Walzer (1983) and Miller (2003) both insist that a theory of justice should hold different principles for distinct sorts of social

relations. In this sense, these results from evolutionary research corroborate a pluralist account of justice.[27]

In the following subsection we will shift our discussion to the findings of experimental economics, a recent and growing field that has been providing interesting insights into the workings of our concept of justice.

(iii) Findings from Experimental Economics

Economists' interest in laboratorial experiments started roughly in the 1940s, with the seminal work of Edward Chamberlin (Davis & Holt, 1993). Chamberlin (1948) designed and implemented a groundbreaking classroom experiment with his students that simulated the workings of the market. He used this artificial market to test neoclassical predictions about price mechanisms. It was the first time that economists availed themselves of a laboratorial experiment so as to test behavioral hypotheses.

Amongst Chamberlin's prominent students was Vernon Smith. Decades later, Vernon Smith was to be awarded the Nobel Prize precisely for his achievements in the field now known as experimental economics. From Chamberlain's first classroom experiments, experimental economics evolved in the direction of the investigation of the behavioral implications of non-cooperative game theory, followed by the development of experiments to test the content of the behavioral axioms from expected utility theory.

In relation to our present subject of interest, namely, justice, experimental economics has provided important evidence on three main fronts: Game Theory, Social Choice, and Behavioral Economics. In the remainder of this subsection, I will respectively go through the main results achieved on each of these fronts.

Game Theory

Laboratory studies of strategic games reveal the widespread existence of "other-regarding" behavior, such as acts of cooperation in the face of material incentive to free ride, or insisting on an equitable share when obtaining one is costly.

[27] It is worth stressing that the validity of this corroboration can be highly disputed. Nevertheless, this debate is beyond the scope of this paper, which merely intends to point out interesting findings and possible contributions for theories of distributive justice.

Among these games, the most popular are the Ultimatum Game (UG) and the Dictator Game (DG). In the UG, there are two players, the Proposer and the Recipient. The goal is to divide an X amount of money between the two players. The division occurs according to the following rules: the Proposer receives the X amount and has to make an offer to the Recipient, who then can either accept or reject it. If the recipient accepts the offer, both receive the designated amount; if he rejects it, both players end up with nothing. The DG is very similar, the difference being that the Recipient has "no voice" in the game; the division proposed by the Proposer will be implemented. The game theoretical equilibria of these games are, respectively: (UL) the Proposer offers the least possible positive amount to the Recipient, and he accepts it; and (DG) the Proposer offers the Recipient zero and keeps the total amount for himself.

Yet these are not the results that experimental economists report on these games. For instance, Kahneman, Knetsch & Thaler (1986) present the findings from an ultimatum game experiment in which the Proposer has to divide ten dollars with the Recipient. Their results show that most Proposers offer five dollars and that offers lower than that amount are often rejected. There is cross-cultural evidence on Ultimatum Games that corroborate these results.

One could interpret these results in different ways. They could mean that we have a "taste" (or preference) for fairness (understood as equal division). But they could also mean that we have the desire to maintain our reputation, which might create expectations capable of altering the Proposer's behavior. With that in mind, Hoffman et al. (1994) propose distinct experimental settings especially designed to affect subjects' expectations about three types of fairness norms: equality, equity, and reciprocity. They take the classical experimental setting discussed above to invoke the norm of equality. On this view, individuals are not differentiated and are just told to divide some amount of money between one another in accordance with the rules. In this context, any deviation from an equal split is likely to be interpreted as cheating on the 'social contract.'

In order to invoke equity, Hoffman *et al*. (1994) explore two interesting variations of the ultimatum game: (i) the exchange treatment (ET), and (ii) the contest treatment (CT). ET assigns the respective roles of seller and buyer to players, and describes the act of dividing an X amount of money as a market transaction, in which the seller (Proposer) chooses a "price" (the share he is willing to divide) and the buyer (Recipient) decides to either buy (accept the offer) or not (turn down the offer). CT

provides each player with a test (general knowledge quiz) so as to allow the participants an opportunity to "earn the right" to play the Proposer's role.

The idea behind these two treatments is to provide a rationale for unequal distributions. Under the exchange treatment, sellers will be perceived as *entitled* to a profit. Alternatively, under the contest treatment, higher achievers will be perceived as *deserving* of a higher 'reward.' This is precisely what is observed in the results of the experiments: a significant move in the direction of the game theoretic equilibrium, with no change in the rejection rates.

Lastly, to invoke the reciprocity norm, Hoffman et al. (1994) designed a "double blind" dictator game for the investigation of the extent to which social isolation influences reciprocal behavior. The idea was to design the experiment so that no one, not even the experimenter, knows which participants are not reciprocating. The hypothesis is that, in such a scenario, individuals will have no worries about reputation and will therefore be more likely to not reciprocate. And that is precisely what happens: under this treatment 64% of the Proposers offer zero to the Recipient, and about 90% of the Proposers offer only one quintile or less of their total designated amount. These results interestingly match the findings from evolutionary psychology on altruistic behavior, as discussed in the previous subsection. As argued by many evolutionary psychologists, we are only interested in being altruistic if this behavior is going to be socially acknowledged. If such is not the case, we will act in a self-interested manner.

Another area of research within the realm of experimental economics is constituted by experiments intended to test the axioms of the so-called social choice theory. This theory predicts how individuals would make choices in society, given some ideal conditions. In the next subdivision we will examine a subset of these experiments, namely the ones designed so as to test how individuals make social choices about the distribution of resources – mainly, income and wealth.

Social Choice

As aforementioned, I will here be exclusively interested in the experiments that investigate social choices over the distribution of resources. In this vein, Herne & Suojanen (2004) developed an experiment that investigates the role of information in decisions about income distributions. The aim of their experiment was to test whether the veil of ignorance hypothesis is crucial in the generation of the Rawlsian principles

of justice. Their results show that under an experimental condition that mimics the deprivation of information specified by the veil of ignorance, the principles most frequently chosen are not Rawlsian, as expected, but a mix of income maximization subject to a floor constraint. Alves & Rossi (1978), Curtis (1979), and Frohlich et al. (1987), and Frohlich & Oppenheimer (1992) have implemented similar experiments that corroborate the choice of a hybrid principle under the veil.

These results fit well with the observation made by Frohlich & Oppenheimer (1996) that a concern for justice does not explain the cooperation between players in impartial prisoner dilemma games. Yet this concern explains cooperative behavior in non-impartial games. The authors claim that, in impartial games, a concern for fairness does not trigger cooperation because there is no conflict between the just allocation and the one that maximizes self-interest.

Along similar lines, Herne & Mard (2006) investigate whether three distinct methods of achieving impartiality are conducive to distinct choices and arguments about the justice of income distributions. The three methods they compare are: (i) the Rawlsian method, in which the impartiality comes from the veil of ignorance; (ii) the method used by Hume and Smith, in which the impartiality is obtained through an impartial spectator (the differences between Hume and Smith are not considered); and (iii) the method developed by Scanlon, in which impartiality is obtained through the use of a device similar to the Rawlsian original position, but without the need for the veil of ignorance. The results showed a larger number of choices of the Rawlsian principle under the second and the third treatments, while under the first treatment the most chosen principle was once again the hybrid of income maximization with a floor constraint.

In the face of the practical impossibility of perfectly reproducing every condition of Rawls' original position, it remains an open question how precisely the veil of ignorance affects the individual decisions. In an attempt to shed more light on this issue, Frohlich et al. (1987) correlated individuals' revealed preferences over income distributions with the following factors: risk aversion, economic status, aspired income level, and political ideology. The results revealed that the only factor that significantly positively correlates with preferences over income distributions is economic status.

In another interesting study, Michelbach *et al.* (2003) designed an experiment in order to be able to synthesize the main theoretical approaches to distributive

justice. In doing so, their aim was to investigate how individuals use distributive principles in judgments concerning income distributions under conditions of strict impartiality. Their subsidiary goal was to discover if the choices made by real individuals under conditions of impartiality are consistent with the predictions of the Rawlsian theory of justice. The results indicate that judgments of distributive justice follow a structure: individuals tend to apply several principles simultaneously (pluralism) and to weight them accordingly to some predictive factors, such as gender and race. A minority of individuals actually utilizes the maximin strategy proposed by Rawls.

These social experiments and their results are still highly contested by political philosophers, and their attitude towards this new field is not illegitimate. Yet it is also important to stress that the findings being revealed by social choice experimenters can illuminate some features of our theories of distributive justice, and therefore should not be so easily dismissed. These results shed light both on our intuitions and on our patterns of behavior, important elements to be taken into account when theorizing about a subject so filled with practical implications as distributive justice.

To conclude this subsection, we now turn to the examination of findings from the field of behavioral economics.

Behavioral Economics

The research project many behavioral economists like Richard Thaler have been working on in the past decades is related to the empirical study of the theoretical hypotheses behind the idea of *homo economicus*. These studies reveal that our individual economic behavior does not adhere to a rational choice theoretical framework – as assumed by mainstream economists. Instead, our economic behavior fits nicely with the so-called dual processing model.[28] That is, the behavioral economic research project has been revealing our biases when making decisions related to our savings plans, house buying, dietary choices, healthcare choices, and other economic related areas of human behavior.

This area of research started with the investigations of bounded rationality in psychology by Amos Tversky and Daniel Kahneman. Yet soon they began

[28] To be explained later in this paper.

collaborating with economists given the impact of their results on the model of rational-agency, foundational for orthodox market functioning mathematical models.

As we now know, our brain has an amazing ability to process and solve problems; nonetheless, this ability is limited. There is a limited amount of information that we can handle – we are not equipped to scrutinize every single decision we make in our lives. Thankfully, our brains developed "shortcuts" that enable us to quickly decide on a variety of issues. These so-called heuristics usually yield very good judgments that allow us to survive through thousands of small decisions that we have to make on an every day basis and of which we may not even be aware.

However, these same helpful heuristics may sometimes give rise to systematic errors – the so-called biases. The point is not that we occasionally make mistakes in our decision-making process. The point is that many of these mistakes are based on the shortcuts that our brains developed to help us, and this feature adds a systematic character to the mistakes that we usually make. This systematic character presents us with the astounding new possibility of predicting human fallibility.

Thaler & Sunstein (2009) rely on this dual process approach to explain why we make systematic mistakes in a wide variety of choice situations. Psychological research has led to the development of a two-system approach to the way we make choices, the way our minds work when we act (Kahneman 2003). Despite divergences in terminology, these two systems can be called the Automatic System (AS) and the Reflective System (RS). The AS is intuitive and automatic, does not involve what we usually understand as 'thinking,' is associated with the oldest parts of the brain (parts we share with other animals), is uncontrolled, effortless, associative, fast, unconscious and skilled. On the other hand, the RS is reflective and rational, deliberate and self-conscious, controlled, effortful, deductive, slow, self- aware, and rule-following.

Kahneman (2003) associates the AS with intuition and the RS with reasoning. He understands both systems along similar lines. As explained above, he understands intuition as spontaneous and effortless, and reasoning as rational, complex, and effortful. Kahneman explains the relations between these two systems: the RS can be said to teach the AS to perform its tasks and also to monitor its performance. The AS, in turn, is responsible for the majority of our thoughts and actions – even if we may not want to admit that.

To better understand this "learning and teaching relation" between the AS and the RS, let us think about how we perform our daily tasks. As already discussed, we

have to do numerous activities – e.g. brushing our teeth, driving, choosing what and where to eat, not to mention the several things we have to do at work. When we learn to drive, for example, it is usually a slow process. In the beginning we have to pay attention to every little detail, but with time we are able to go to work without even realizing that we did that. This is so because at first we have to use our RS to learn how to drive. Once the RS learned it well and performed it many times, it can delegate the repeated task to the AS, so that we can use our RS to think about other important issues. It does not mean that the RS can teach the AS to perform all kinds of tasks – we still have to use the RS to solve different problems and to deliberately reason about any subject. Thus, the idea is that when we are exposed to repetitive tasks the RS can first learn to execute it and, with practice, "teach" the AS to do it by itself.

Psychologists and neuroscientists argue that we have this dual system because the most developed part of our brain, the neocortex, is not able to carry out all the activities demanded from us on a daily basis. Therefore, we tend to use our RS only when confronted with problems that require active reasoning. Still, Kahneman claims that the RS monitors the AS, correcting its decisions whenever possible.

Kahneman and Frederick (2002) describe this monitoring process in the following way: "System 1 quickly proposes intuitive answers to judgment problems as they arise, and System 2 monitors the quality of these proposals, which it may endorse, correct or override"(2002, 51); emphasizing that "errors and biases only occur when both systems fail"6(2002, 52). However, the monitoring that the RS executes is usually loose, consequently allowing many erroneous judgments reached by the AS to be expressed in human action. Ellen J. Langer (1992) refers to these erroneous judgments as 'mindless behavior.'

Psychologists and behavioral economists have documented several biases in the past few decades.[29] Amongst the main biases documented so far, one finds the following: (i) *Illusion of Validity* – people have a tendency to be overconfident in their own judgments, even in the light of evidence that their judgments are wrong; (ii) *Anchoring Bias* – whenever people are exposed to a number or reference-point, their judgment is influenced by that reference whether they intended to be influenced or not; (iii) *Status Quo Bias* – people have a tendency not to bother to opt out of default

[29] See: Kahneman and Tversky (1973; 1979), Kahneman et al. (1982), Kahneman et al. (1990; 1991), and Tversky and Kahneman (1974; 1981; 1986; 1991; 1992).

rules; (iv) *Endowment Effect* – people tend to overvalue what is already in their possession; (v) *Framing Effect* – when confronted with a set of alternatives, people's choices are influenced by the manner in which the set is arranged; (vi) *Projection Bias* – people tend to project their current emotional state into the future; (vii) *Availability* – people tend to be more aware about risks readily available than to those they rarely hear about; (viii) *Benefits Now, Costs Later* – people tend to avoid present costs and to seek present benefits; and (ix) *Follow the Herd* – people have a tendency to conform to other's behaviors.

Richard Thaler and Cass Sunstein have recently applied this knowledge about our supposedly irrational behavior to public policy, in their book entitled *Nudge* (2008). Their main point is that we currently know both how these biases work and that they are systematic. Additionally, we have enough information so as to be able to predict under which circumstances they are most likely to arise. Consequently, we are capable of influencing people's behaviors by changing those circumstances – i.e., by changing the choice architecture. This is what *Nudge* is about: altering the circumstances in which people find themselves making choices, with the purpose of influencing their behavior in some desired direction.

In the next subsection, I will transition to a related area of research, namely, moral psychology. This area is related to experimental economics in the sense that its focus is also on human behavior. The difference relies in the fact that moral psychologists are solely interested in the moral domain, while experimental economists share a broader interest.

(iv) Findings from Moral Psychology

There is a large body of research in moral psychology (and more recently in experimental philosophy) about the nature of our moral intuitions. A considerable part of this literature is related to the aforementioned dual process model of the mind.[30] The paradigm of this research project can be understood along the following lines: our moral intuitions are more frequently than not affected by a variety of factors, such as emotional states, environmental cleanliness, odors, order, etc. These factors are in the majority of cases irrelevant from the perspective of morality.

[30] See preceding *Subsection*.

For instance, Unger (1996) contends that our intuitions about morality are subject to order effects. That is, our moral judgment shifts if two alternatives are presented either as a pair or as part of a list with additional alternatives. Kahneman & Tversky (1979) provide another example of such framing effect in an experiment where the participants were faced with two equivalent disease prevention programs, and they had to choose which was morally preferable. Both programs yielded the exact same expected results. Yet they were presented to participants under different framings, all in terms of probabilities. While one program emphasized the number of people who would likely be saved, the other program emphasized the number of people who would likely die. Instead of realizing that the results were the same and being indifferent to which program would be implemented, people always chose the one that expressed less expected deaths and more expected saved lives. Just as another example, Wheatley & Haidt (2005) showed that moral judgments are also affected by disgust. In the experiment, when a story contained the word that elicited disgust participants were more likely to strongly morally condemn the acts.

A related line of research in moral psychology investigates the role of reason in our moral judgments. Haidt, Koller & Dias (1993), Haidt (2001), and Haidt & Hersh (2001) provide cross-cultural evidence that when people are confronted with harmless but supposedly offensive actions, or when they are questioned about issues related to sexual morality, they are usually 'morally dumbfounded.' That is, people "stutter, laugh, and express surprise at their inability to find supporting reasons, yet they would not change their initial judgment of condemnation" (Haidt, 2001, p. 346). In their experiments, they ask individuals to judge the wrongness of a variety of moral dilemmas. After they have decided which acts are morally right or wrong, the individuals are given a series of arguments that invalidate all of the tentative reasons they try to provide in support of their moral judgments. The interesting result is that they end up admitting that they do not have any reason to hold that particular judgment, yet they do not change their minds – they just know it is wrong, without knowing why.

Haidt interprets these results as evidence that rationalist approaches[31] to morality are not appropriate. Based on these and other results from several

[31] Rationalist approaches state that moral judgments are primarily reached through reasoning and reflection, i.e., through the use of our distinctively human rational capacity. Additionally, in rationalist models, moral emotions are never the direct cause of moral judgments.

experiments (2001, 2003), he proposes an alternative model to explain how moral judgments are arrived at: the so-called *Social Intuitionist Model*. The Social Intuitionist Model explains moral judgments as the result of both social relations and intuitions.[32] Regarding the latter, moral judgments are interpreted as the result not of a rational process but rather of a process more similar to perception; one "just sees without argument that they are and must be true" (Harrison, 1967, p.72). As to the former, Haidt argues that moral judgments should be investigated as an interpersonal process. He goes on to claim that "moral reasoning is usually an ex post facto process used to influence the intuitions (and hence judgments) of other people" (Haidt, 2001, p. 344).

The empirical findings reported by Haidt provide four reasons for questioning the rationalist model of morality: (i) moral judgments involve the use of both reasoning and intuitive cognitive processes, and the former has been overemphasized; (ii) human reasoning appears to be always motivated; (iii) our experience of objective reasoning is illusory, justifications are usually constructed by *post hoc* reasoning; and (iv) there is a higher degree of covariance between moral action and moral emotion than between moral action and moral reasoning.

Haidt argues that the Social Intuitionist Model is capable of better accounting for the way in which people actually arrive at moral judgments. Most importantly, moral judgments are rooted in our "hot" affective system, not in our "cold" rational abilities. In his words, "Reason can let us infer that a particular action will lead to the death of many innocent people, but unless we *care* about those people, unless we have some *sentiment* that values human life, reason alone cannot advise against taking the action"[33] (Haidt, 2001, p. 345).

In addition, Haidt claims that people access their a priori moral theories in order to provide socially acceptable reasons for praise and blame. As he stresses, people consult "a pool of culturally supplied norms for evaluating and criticizing the behavior of others" (Haidt, 2001, p. 352). Hence the idea of *post hoc* moral reasoning based on culture.

Regarding our moral emotions, there is to date no complete general taxonomy, but moral psychologists have been advancing in the field. Haidt (2003) provides a

[32] "Moral intuition is a kind of cognition, but it is not a kind of reasoning." (Haidt, 2001, p. 344)
[33] This quote can be taken as an argument against the Liberationist view – to be discussed in the next section.

very useful categorization that divides emotions into two main sets: (i) self-conscious, and (ii) other-conscious. The self-conscious emotions are related to self-assessments and arise from a concern about the opinions of others on self-behavior and self-identity. Amongst these one can find two subsets: the self-critical and the self-praising emotions. The former are related to reduced social status or self-esteem, and comprise the emotions of guilt, shame, and embarrassment. The latter are related to increased social rank and self-esteem, and comprise the emotion of pride.

The other-conscious set regards the others-directed emotions. These again are divided into the following subsets: other-critical, other-praising, and other-suffering emotions. The first subset consists of emotions responsible for the punishment of the violators of social rules, and includes indignation, anger, contempt, and disgust. The second subset contains the emotions that drive reciprocity and cooperation, consisting of gratitude and awe. The third and final subset embraces the emotions that are central in our helping and altruistic behaviors, which are pity and compassion.

The take home lesson from these empirical findings reported by moral psychologists is nicely spelled out in the following passage:

> Kant has had a much larger impact than Hume on modern moral philosophers (e.g., R.M. Hare, 1981; Rawls, 1971), many of whom have followed Kant in attempting to deduce a foundation for ethics from the meaning of rationality itself. (…) Rather than following the ancient Greeks in worshipping reason, we should instead look for the roots of human intelligence, rationality, and virtue in what the mind does best: perception, intuition, and other mental operations that are quick, effortless, and generally quite accurate.
> (Haidt, 2001, pp. 345; 351)

Hence the findings from moral psychology point in the direction of a human morality highly influenced by our emotions. In this sense, these results add support for a Humean approach to justice, and fuel the debate between rationalists and sentimentalists about morality.

In the next subsection I will discuss the findings from two fields that have to date not influenced the political philosophical debate about justice in any manner. Yet, as we will be able to appreciate, these findings can pave the road for clarifying the reasons behind political disagreements.

(v) Findings from Social and Political Psychology

Social scientists have been providing accumulating evidence regarding the human tendency to maintain existing social arrangements. In the face of these data, social and political psychologists have been interested in understanding the psychological underpinnings of such behavior. The most promising approach that has been guiding contemporary research in this area is the so-called *Motivated Social-Cognitive Approach*. This approach builds on the theories that link social and cognitive motives and processes to social content, interpreting the support for specific ideologies as a means to the satisfaction of diverse psychological needs. The aim is to convey a unified account of social ideologies that embraces simultaneously social, cognitive and motivational factors (Jost, Banaji & Nosek, 2004).

The motivated social-cognitive approach works at three distinct levels. Some ideologies are supported at the individual level, some at the group level, and yet some work at the systemic level. At the individual level, the idea is that persons have a tendency to do cognitive and ideological work as a consequence of specific individual psychological features or in order to satisfy their existential and epistemological needs. At the group level (Group Justification Theories – GJT), this work is undertaken so as to justify group identity. And at the systemic level (System Justification Theories – SJT), this psychological work is done in order to justify the existing broader social arrangements. In the following, I will address the theories under each one of these support levels, respectively.

(A) The Individual Level

On this level, political psychologists have been trying to understand, for instance, the endorsement of particular ideologies, such as political conservatism, liberalism or, more recently, libertarianism. They use ego-justifying theories to comprehend the embracement of these particular ideologies; mainly, *theories of personality* and *theories of epistemic and existential needs.*

Theories of Personality

These theories include amongst its principal conceptions (a) right wing authoritarianism; (b) intolerance of ambiguity; (c) mental rigidity, dogmatism and

closed-mindedness; (d) ideo-affective polarity; and (e) uncertainty avoidance. Each of these conceptions will be briefly detailed in what follows.

(a) Right Wing Authoritarianism (RWA)

This conception was first developed by Adorno and the intellectual successors of the Frankfurt School, and subsequently improved by Altemeyer (1981), who defined RWA as submission to established and legitimate authorities, sanctioned general aggressiveness towards various persons, and adherence to the generally endorsed social conventions.

(b) Intolerance of Ambiguity

This conception, first elaborated by Frenkel-Brunswik (1948), argues for intolerance of ambiguity as a general personality variable that positively correlates with prejudice and other social and cognitive variables. Empirical evidence supports the existence of a positive correlation between this personality trait and cognitive and motivational tendencies to seek certainty and cling to the familiar, reach premature conclusions, and impose clichés and stereotypes.

(c) Mental Rigidity, Dogmatism and Closed-mindedness

This conception was developed by Rokeach (1960) to address the recurrent criticism directed at Adorno's work on authoritarianism, which was taken to explain the presence of authoritarianism only amongst right-wingers. The idea is that these broader cognitive-motivational factors (mental rigidity, dogmatism and closed-mindedness) constitute the general framework under which authoritarian personalities tend to function.

(d) Ideo-affective Polarity

This conception was developed by Tomkin (1963) in order to grasp the role of affection and motivation in social ideologies. It constitutes a groundbreaking approach due to its interpretation of ideological predilections as permeating every domain of a person's life – such as attitudes towards the arts, music, science, and so on. According to Tomkin (1963, 1965) the ideological left is associated with liberty and humanism, while the ideological right is associated

with rule following and normative concerns. The theory is illuminating in regard to its assessment of the affective and motivational basis of conservatism (anger, contempt, and the desire for punitiveness), and to its suggestion that conservatives are motivated to follow rules in a wide variety of domains not restricted to the political universe.

(e) Uncertainty Avoidance

This conception was developed by Wilson (1973) and provides an interpretation of the motivational core of all the attitudes embraced by conservatives. This motivational core is characterized by a generalized susceptibility to experience threat or anxiety in the face of uncertainty.

Theories of epistemic and existential needs

These theories are based on specific cognitive and motivational human needs to know, and to cope with existential anxiety. Amongst its principal conceptions are (a) lay epistemic theory; (b) regulatory focus theory; and (c) terror management theory. Each of these conceptions will be briefly detailed in what follows.

(a) Lay Epistemic Theory

This conception focuses on the epistemic need for cognitive closure as a personality feature related to the embracement of specific social ideologies. Under this interpretation, persons with differing needs for closure are not indifferent to ideological content. For instance, a higher need for closure tends to be positively associated with conservative attitudes.

(b) Regulatory Focus Theory

This conception provides a reading of our desired goals as subdivided into two major systems: the Promotion System, related to hopes and aspirations, and aimed at accomplishment; and the Prevention System, related to duties and obligations, and aimed at safety. To the extent that differing ideologies are psychologically motivated by different sorts of desires, situations that induce the use of one of these systems rather than the other have been shown to induce ideological shifts in the general population.

(c) Terror Management Theory

This conception builds on the relations amongst social ideologies and the human motivations to cope with mortality and existential anxiety. Terror management theory claims that specific ideological attitudes are the consequence of worldview-enhancing cognitions induced by the necessity to shield anxiety-inducing thoughts.

(B) The Group Level

On the group level, *Group Justification Theories* hold that people are driven by ethnocentric motives to build in-group solidarity and to defend and justify the interests and identities of fellow in-group members against those of out-group members. The core characteristics of these theories encompass the view that groups serve their own interests, develop ideologies to justify those interests, have strong preferences for members of their own kind, are hostile and prejudicial toward outsiders, and are conflict-seeking whenever it helps to advance their partisan interests and particularistic identities. According to Jost, Banaji, & Nosek (2004) group justification theories hold the following assumptions:

> Similar others are preferred to dissimilar others. (Allen & Wilder, 1975; Brewer, 1979; Tsui, Egan, & O'Reilly, 1992)
>
> Prejudice is a form of hostility directed at outgroup members. (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950; Allport, 1954; Brown, 2000b; Pettigrew, 1982)
>
> Intergroup relations in society are inherently competitive and conflict- ridden. (Bobo, 1988; Sherif, 1967; Sidanius & Pratto, 1999)
>
> Intergroup behavior is driven primarily by ethnocentrism and ingroup favoritism. (Brewer & Campbell, 1976; Brewer & Miller, 1996; Sumner, 1906; Tajfel & Turner, 1986)
>
> Prejudice, discrimination, and institutionalized oppression are inevitable outcomes of intergroup relations. (Sidanius & Pratto, 1993)
>
> Members of dominant groups strive to impose their hegemonic will on members of subordinated groups. (Fiske, 1993; Sidanius & Pratto, 1999)
>
> Members of subordinated groups first seek to escape the implications of group membership by exercising individual exit and mobility options. (Ellemers, Wilke, & van Knippenberg, 1993; Hirschman, 1970; Tajfel, 1975)
>
> When individual exit/mobility is impossible, members of subordinated groups engage in identity enhancement strategies of resistance and competition. (Scott, 1990; Spears, Jetten, & Doosje, 2001; Tajfel & Turner, 1986)
>
> In coping with chronically threatened social identities, members of subordinated groups typically express stronger levels of ingroup favoritism than do members of dominant groups. (Leach, Spears, Branscombe, & Doosje, 2003; Mullen, Brown, & Smith, 1992)
>
> Political ideology mirrors/group membership individual and collective self-interest and/or social position. (Centers, 1949; Downs, 1957; Olson, 1971; Sidanius, Singh, Hetts, & Federico, 2000).
>
> (pp. 882-883)

Hence for advocates of group justification theories it is as if the advantaged are relentlessly looking to cash in on their dominance and the disadvantaged are proud revolutionaries-in-waiting. Both types of groups are seen as primarily self-interested, and overt conflicts of interest are assumed to be endemic.

(C) The System Level

Both ego-justifying and group justification theories have been recently supplemented by a third level approach, namely, the System Justification Theories. This supplementation has been envisioned as a response to the existence of a particularly vexing, but consistent, social psychological finding: the prevalence of out-group favoritism among low-status group members (Jost & Banaji, 1994).

The SJT examines the process by which existing social arrangements are legitimized, even at the expense of personal and group interest. This theory addresses the antecedents, contents, and consequences of thoughts, feelings, and behaviors that serve to maintain the societal status quo. According to system justification theory, there is a general social psychological tendency to rationalize the status quo, that is, to see it as good, fair, legitimate, and desirable – the classical dispositional outlook of Voltaire's famous character, Dr. Pangloss, who believed that he was 'living in the best of all possible worlds.'

Whether because of discrimination on the basis of race, ethnicity, religion, social class, gender, or sexual orientation, or because of policies and programs that privilege some at the expense of others, or even because of historical accidents, genetic disparities, or the fickleness of fate, certain social systems serve the interests of some stakeholders better than others. Yet evidence shows that most of the time the majority of people – regardless of their own social class or position – accept and even defend the legitimacy of their social and economic systems and manage to maintain a 'belief in a just world.'

Knowing how easy it is for people to adapt and rationalize the way things are makes it easier, for instance, to understand why the apartheid system in South Africa lasted for 46 years, the institution of slavery survived for more than 400 years in Europe and the Americas, and the Indian Caste system has been maintained for 3,000 years and counting. The remaining question is: how do people rationalize bad outcomes for themselves and others and, above all, the social systems that dictate these bad outcomes? Social and political psychologists have been recently uncovering the mechanisms that work to provide these rationalizations, such as: Complementary Stereotypes, Distorted Social Judgments, Rationalization of Likely Outcomes, Belief in a Just World, and Economic System Justification. These mechanisms will be respectively discussed in what follows.

Complementary Stereotypes

Complementary stereotypes are stereotypes that appear to compensate for intergroup disparities by assigning offsetting advantages and disadvantages to low- and high-status groups, respectively. The guiding thesis is that complementary stereotypes serve to rationalize inequality, allowing people to maintain their belief that the societal status quo is, generally speaking, fair, legitimate, and justified. One

prominent example of a complementary stereotype is the idea that poor people are humble and honest, while rich people are greedy and dishonest.

These representations communicate that 'no one group has it all' and thus encourage the feeling that things somehow balance out in a way that makes the system seem fair, or at least not unbearably unfair. Thus, if equality cannot be achieved in actuality, complementary stereotypes may help us to create a comforting illusion of equality.

Distorted Social Judgments

Another manifestation of SJT is the pervasive tendency to use social judgments to justify arbitrary status and power differences between groups. For instance, Haines & Jost (2000) report findings that show that group members arbitrarily (and even illegitimately) ordained with high levels of relative power in an experimental settings tend to be perceived as more intelligent and responsible than others in position of low power. Interestingly, these perceptions are equally shared by the powerful as well as the powerless. From a SJT perspective, the justification of arbitrary inequalities is an important instance of "buying into" the status quo.

Rationalization of Likely Outcomes

Anticipatory rationalization of likely outcomes represents another mechanism through which individuals are capable of justifying the social system to which they belong. To the extent that people are motivated to justify the status quo, they begin to see highly probable events in increasingly favorable terms and highly improbable events in increasingly unfavorable terms. These desirability adjustments take two forms: (i) the "sour grapes" rationalization, and (ii) the "sweet lemons" rationalization. The former refers to the tendency of thinking about good and desirable outcomes that are beyond our reach as bad and undesirable; and the latter refers to the tendency of thinking about bad and undesirable yet highly likely outcomes as actually good and desirable.

Belief in a Just World

Lerner and Miller (1978) report the existence of a universal human need to believe that outcomes are fair and just and that people get what they deserve and deserve what they get. The basic argument is that living in an unpredictable,

uncontrollable, and capriciously unjust world would be unbearably threatening, and so we cling defensively to the illusion that the world is a just place. The origin of the *just world conception* can be traced back to the original empirical findings of Lerner & Simmons (1966). These findings suggest that persons have a tendency to blame the victims of misfortunes for their own fate, a documented tendency known as *derogation effect* – a tendency to see consistency between outcomes and virtue. Some types of individuals are more likely than others to derogate an innocent victim. People with strong religious convictions, for example, appear to derogate less than nonreligious people (Sorrentino & Hardy, 1974).

Based on these empirical findings, Lerner (1965) formulated the Just World Hypothesis. This hypothesis states that individuals have a need to believe that they live in a world where people generally get what they deserve. The belief that the world is just enables individuals to confront their physical and social environment. This belief enables individuals to perceive the world as if it were stable and orderly. The justness of others' fates thus has clear implications for the future of the individual's own fate. As a consequence of the perceived interdependence between their own fate and the fate of others in their environment, individuals confronted with an injustice generally will be motivated to restore justice. To witness and admit to injustices in other environments does not threaten people very much because these events have little relevance for their own fates. As events become closer to their world, however, the concern over injustices increases greatly, as does the need to explain or make sense of the events.

Comer & Laird (1975) provide an interesting experiment in this respect, the so-called 'eat a worm' experiment. They report that people will often alter their conceptual system, in this case their perception of their own worth, to impose order and justice on random events in their lives. Perhaps the most alarming finding to emerge from the study is that most of those who engaged in self-derogation as a consequence of a negative expectation chose to follow through with the negative event even when it was avoidable – thus, it appears that individuals not only change their conceptions of their own worth, but they also actually *act* on these new conceptions.

Economic System Justification

This is understood as a tendency to perceive the existing social, economic and political arrangements as inherently fair, legitimate and justifiable. Advocates of system justification theory claim that the justification of the status quo is delivered by processes that work at the expense of personal and group interests; but in the best interest of the maintenance of the existing social system. One example of these processes is the adherence to unfavorable stereotypes undertaken by disadvantaged groups. For instance, evidence shows that African American respondents generally accept self-derogating stereotypes as lazy, irresponsible, and violent (Piazza, 1993).

Within the economic system justification framework, Jost *et al*. (2003) have determined a more specific type of system justification: the Fair Market Ideology. They argue that institutional entities like the free market system and specific institutionalized practices survive, at least in part, because people accept them as legitimate and therefore protect and sustain them over time. Perceptions of legitimacy, in turn, depend at least in part upon ideological factors. For instance, one's ideological beliefs, values and goals affect the likelihood of judging existing institutional forms and practices to be fair, legitimate and just and therefore deserving of continued support.

Ideologies are complex belief systems that incorporate, among other things, people's theories about human nature, their philosophies concerning the appropriate use of social power, status and authority, and their moral and pragmatic convictions concerning the maximization of social and economic welfare. There are demonstrable links between ideological orientations and preferences for specific justice principles, such as liberalism and equality, on the one hand, and conservatism and equity, on the other.

In this context, the fair market ideology (FMI) is defined as the tendency to view market-based processes and outcomes not simply as efficient, but as inherently fair, legitimate and just. This ideology would account for the continuous perception of the current economic system as fair and legitimate, despite the unprecedented increases in wage dispersion over the last twenty years. It would also account for an even more striking empirical finding – namely, the relatively large number of self-designated have-nots who accept the fairness and legitimacy of the economic system.

Jost *et al*. (2003) report that people who are especially prone to endorse the fair market ideology are also more likely to believe in a just world, to engage in self-deception, to accept power distance, to oppose equality, and to be politically conservative and even authoritarian. Additionally, according to this ideology, selfishness is not only understood as rational; given that it conforms to the underlying assumptions of a market-based system, it is perceived as actually fair.

Hence, as discussed above, we are able to distinguish among three different justification tendencies or motives that have the potential to be in conflict or contradiction with one another for members of disadvantaged groups (Jost & Banaji, 2004): ego justification, group justification, and system justification. Within this theoretical framework, one can see that members of disadvantaged groups are likely to engage in social change only when ego justification and/or group justification motives overcome the strength of system justification needs and tendencies – therefore our unfortunate tendency to perpetuate inequalities and injustice.

In a distinct yet related strand of research, political psychologists have been investigating the morality underlying our divergent political beliefs: ranging from liberalism to libertarianism. Within this strand, Haidt and colleagues have an active research agenda that has been illuminating the morality of our political convictions. The goal is to better comprehend the difficulties (apparently inherent) involved in the achievement of consensus in the political debates.

Richard Shweder (1990) has provided evidence for the existence of three dimensions of morality: (i) the ethics of autonomy, e.g. rights, justice, fairness, and freedom are moral goods because they help to maximize the autonomy of individuals, and to protect individuals from harms perpetrated by authorities and by other individuals; (ii) the ethics of community, e.g. key virtues are duty, respect, loyalty, and interdependence; and (iii) the ethics of divinity, e.g. the body is viewed as a temple housing divinity within, thus moral regulations should help people to control themselves and avoid sin and spiritual pollution in matter related to sexuality, food, and religious law more generally.

Haidt & Joseph (2004) build on Shweder's theory by adding two additional dimensions that correlate with his framework. They argue for the existence of the following five foundations for all moral systems:[34]

(a) Harm: This foundation is related to our long evolution as mammals with attachment systems and an ability to feel (and dislike) the pain of others. It underlies virtues of kindness, gentleness, and nurturance.

(b) Reciprocity: This foundation is related to the evolutionary process of reciprocal altruism. It generates ideas of justice, rights, and autonomy.

(c) Ingroup: This foundation is related to our long history as tribal creatures able to form shifting coalitions. It underlies virtues of patriotism and self-sacrifice for the group. It is active anytime people feel that it's "one for all, and all for one."

(d) Hierarchy: This foundation was shaped by our long primate history of hierarchical social interactions. It underlies virtues of leadership and followership, including deference to legitimate authority and respect for traditions.

(e) Purity: This foundation was shaped by the psychology of disgust and contamination. It underlies religious notions of striving to live in an elevated, less carnal, more noble way. It underlies the widespread idea that the body is a temple which can be desecrated by immoral activities and contaminants (an idea not unique to religious traditions).

Each foundation is taken to have its own evolutionary history that gives rise to moral intuitions across countries. In addition, each dimension is akin to a kind of 'taste bud', producing affective reactions of liking or disliking when certain kinds of patterns are perceived in the social world. Cultures are understood as varying in the degree to which they construct, value, and teach virtues based on these five intuitive foundations.

More recently, Haidt and colleagues have been applying these findings and theories to study the nature of political dissent. Their motivating research question is: why is it so hard to achieve consensus in the political sphere? They ask this question in the background of American political split between republicans (a curious mix of conservatives and libertarians) and democrats (liberals) and their increasing inability to engage in any sort of meaningful dialogue. This novel research program has already generated interesting findings. For instance, Haidt and colleagues report evidence showing that there is a moral distinction between liberals, conservatives, and

[34] As in: www.moralfoundations.org

libertarians. In terms of Haidt's moral foundations, they endorse, respectively, the first three, all five, and solely the first. These differential foundations seemingly help to explain their baffling inability to fruitfully debate political issues.

Of all the subsections in which I report the recent empirical research on the conception of justice, this is certainly the one that addresses the most neglected data by political philosophers. It is beyond the scope of this paper to unveil the reasons for this current state of affairs. What is interesting about it is that given the fact that this literature has been widely overlooked it may comprise surprising implications for contemporary theories of distributive justice. In this sense, it is worth investigating which sorts of implications may be entailed by political psychology.[35]

(vi) Findings from Neuroscience

The past few decades have witnessed the birth and growth of a remarkably new experimental agenda: the study of the neural underpinnings of human morality (Moll *et al*., p. 2, 2008). This new agenda has been responsible for adding scientific support for the view that our emotions play a primary role in the generation of moral judgments. For instance, Moll *et al*. (2001) and Moll *et al*. (2002) have provided cumulative evidence that the emotional parts of the brain are the more activated ones when people think about sentences with moral content. Moreover, Damasio *et al*. (1990) and Damasio (1994) have showed that brain damage can impair the development of cognitive-affective connections, consequently impairing the development of moral competence.

Economists have also been doing research with functional neuroimaging themselves. One example of such interdisciplinary field is the analysis of brain imaging during ultimatum games. The idea is to observe which areas of the brain are more active when individuals engage in distinct strategies. As already discussed, in ultimatum games the proposer formulates a proposal for sharing money that can be seen either as fair or unfair by the recipient who, in turn, responds accepting of rejecting it. Sanfey, Rilling, Aronson, Nystrom & Cohen (2003) report findings that show that unfair offers produce increased activity in the anterior insula, an area associated with anger, disgust, and autonomic arousal.

[35] Needless to say, this is a mission to be undertaken in future works.

In a distinct but related vein, Greene (2001, 2008) and Greene *et al*. (2004) report extensive empirical evidence in favor of a so-called *dualistic* view of moral reasoning. Greene claims that our moral judgments can be (roughly) divided into two very broad categories. On the one hand, we have deontological judgments that encompass moral rules valued in and of themselves, such as rights and duties. On the other hand, we have consequentialist judgments that attach moral value to a cost-benefit analysis of the consequences of human acts.

Greene gathers evidence from neuroimaging involving a series of moral dilemmas in which the participants have two morally unsound options to choose from, being one representative of consequentialist and the other of deontological morality. The dilemmas are variations of the trolley problem, including the footbridge and the loop case versions.[36] He reports that whenever participants opt for the deontological option, the brain regions activated are the ones responsible for emotions; and whenever the consequentialist alternative is chosen, the brain regions activated are related with rational cognition. Greene (2008) interprets these findings as revealing of the psychological rationale that underlies the main moral philosophical theories.

In this vein, Greene contends that consequentialist and deontological views of philosophy are not so much philosophical inventions as they are philosophical manifestations of two psychological patterns. Much to the discontentment of deontological philosophers, Greene's findings have been suggesting that our core deontological judgments are no more than emotionally driven moral judgments. He argues that when we explore the psychological causes of characteristically deontological judgments we end up finding that deontological moral philosophy is ultimately an attempt to produce rational *post hoc* justifications for our affectively generated moral intuitions (Greene, 2008, p. 39).

Moll *et al.* (2008) present evidence about the origin of moral judgments that contrasts with the dualistic view defended by Greene *et al.* (2004). According to the observations of neural activity revealed by Moll *et al.* (2008), moral emotions are not in competition with rational processes during moral judgments. Hence, they embrace a complementary view, in which emotion and rational cognition work together in the generation of moral judgments. As the authors stress, "Most likely, moral emotions

[36] References for trolley problems: Foot (1978), Thomson (1976, 1985).

help guide moral judgments by attaching value to whichever behavioral options are contemplated during the tackling of a moral dilemma" (Moll *et al.*, 2008, p. 5).

Much earlier, Gazzaniga, Bogen & Sperry (1962) present empirical findings that help support the hypothesis of moral rationalization elaborated by Haidt, Greene, and others. The idea, as already mentioned, is that our moral judgments have only the appearance of rationality; in fact, they are not generated by reasons, but by affective processes that are followed by the elaboration of *post hoc* reasoning. Gazzaniga *et al.* report findings about split-brain patients that show this *post hoc* reasoning, which they call moral confabulation:

> Split-brain patients show this effect in its most dramatic form. When the left hand, guided by the right brain, performs an action, the verbal centers in the left brain readily make up stories to explain it (Gazzaniga, Bogen & Sperry, 1962). The language centers are so skilled at making up post hoc causal explanations that Gazzaniga (1985) speaks of an interpreter module. He argues that behavior is usually produced by mental modules to which consciousness has no access but that the interpreter module provides a running commentary anyway, constantly generating hypotheses to explain why the self might have performed any particular behavior. (Haidt, 2001, p. 352)

In face of all these novel findings, Moll *et al.* (2008) argue for the relevance of a free exchange of ideas between neuroscience and moral philosophy. They stress that "moral emotions might prove to be a key venue for understanding how phylogenetically old neural systems, such as the limbic system, were integrated with brain regions more recently shaped by evolution, such as the anterior PFC, to produce moral judgment, reasoning, and behavior" (Moll *et al*., 2008, p. 17). Yet the pace of development of the field of moral neuroscience will critically depend on open and unbiased scientific discussions, and on the design of experiments and models in which the humanities and the biological sciences work together.

Contributing to this new interdisciplinary project, philosopher Richard Joyce has engaged in the philosophical analysis of the above discussed results. He argues that recent research from neuroscience does not support emotivism (Joyce, 2008) – contrary to what many investigators have been stating. For instance, Greene and Haidt interpret the recent findings as suggesting that our morality is more a matter of emotion and affectively charged intuitions than of deliberative reasoning. To this view Joyce replies that surely we might think this way, if we are willing to say that

just because "when we hook up people's brains to a neuroimaging device, get them to think about moral matters, and observe the presence of emotional activity, emotivism is supported." (p. 375) Joyce goes on to state that:

> (…) the most that neuroscientific discoveries could establish is that public moral judgments are accompanied by emotions, and perhaps that they are caused by emotions, but further arguments would be needed to show that public moral judgments *express* those emotions. It is entirely possible that moral judgments are typically caused by emotional activity but nevertheless function linguistically as assertions (i.e., expressions of belief). (p. 375)

What about moral rationalism? Does the neuroscientific empirical evidence threaten it? It depends on what researchers mean when they refer to 'moral rationalism'. Joyce (2008) argues that some kinds of rationalism are indeed threatened by the recent findings, but some other types remain immune. One would first have to clarify the concept, drawing the appropriate distinctions, so that one could then see which kinds of rationalism remain unchallenged by the empirical work.[37]

3. Implications for Political Philosophy: roads for future research

All of the novel and exciting empirical findings brought to light in the previous section present political philosophy with a number of important implications and possible roads for future research. My goal in this last section is to highlight at least a few of these implications, pointing to the importance of paying due attention to the relevant empirical sciences when developing theories of justice.[38]

For starters, one important insight provided by the natural sciences concerns the origin of our moral systems. They illuminate how moral rules emerged in primate and human societies, helping us to better understand what morality is and how it came about. For instance, one of the central problems of societies, according to John Rawls (1971, p. 4), is the resolution of two seemingly undeniable and incompatible facts about social reality. On the one hand, individuals are not indifferent to the distribution of the fruits of joined labor – the so-called conflict of interests. On the other hand,

[37] Vide third paper.
[38] The arguments for an empirically informed political philosophy were fully developed in the first paper.

individuals agree that a cooperative enterprise makes it possible for every member of society to enjoy a better life – the so-called identity of interests.

Evidence gathered by primatologists beautifully demonstrates that both facts are indeed undeniable, insofar as they are natural, and surely compatible (at least from a biological point of view), given its pervasive existence in primate societies. Fleck & de Waal (2002) present an interesting insight regarding this point, stating that all evidence points to a morality that was not devised to subjugate our independent interests. Rather, a moral system emerged precisely out of the interaction of the two sets of interests – collective and individual. They argue that human morality is best understood as having arisen out of an implicit agreement among group members that enabled individuals to profit from the benefits of cooperative sociality.

Results from evolutionary biology point in the same direction when it comes to the origins of human morality. That is, evolutionary biologists claim that our moral rules emerged as a solution to a cooperative game problem played by self-interested individuals – the aforementioned problem described by Rawls. Evolutionary biologists, psychologists, and even philosophers are able to show how altruistic behavior emerged using the tools provided by dynamic modeling of social learning.

There is a striking similarity between this mode of proceeding in the study of moral norms and the Humean approach portrayed in his *Treatise of Human Nature*. Hume was temporally deprived of primatology and evolutionary theory; nonetheless, the story he tells about the origin of justice holds a remarkable resemblance to modern evolutionary game theoretical explanations of morality. This is revealing of the importance for political philosophers to have a better comprehension of the genealogy of our moral rules. Brian Skyrms (2002) points out that, just like evolutionary theorists, "Hume is interested in how we actually got the contract we now have. He believes that we should study the processes that lead to a gradual establishment of social norms and conventions," and goes on to add that "The proper way to pursue modern Humean social philosophy is via dynamic modeling of cultural evolution and social learning" (p. 272). Even if one may disagree with Skyrms about the proper way to engage in Humean philosophy, political philosophers should nevertheless pay due attention to all these empirical evidence we now have easily available.

The empirical sciences have also been illuminating, as previously discussed, the nature of our moral decisions. And here once again we are impelled in the direction of Hume's understanding of morality: empirical results arising from an array

of distinct disciplines are revealing our moral decisions to be more related to our sentiments than to our purely rational capacities. For instance, Fleck & de Waal (2002) claim that

> The primate research implicitly suggests that this emphasis on the role of emotions is both insightful and accurate – in primate groups individuals are motivated to respond to others based on the emotional reactions they have to one another's behavior.
> (p. 20)

This suggestion does not amount to an exclusively emotional view of human morality. It would be erroneous to equate moral emotions with lack of rationality and judgment. The emotions Fleck & de Waal claim to be involved in morality are very complex and require the use of reasoning. Put it even more sharply,

> Perhaps primate research that suggests that morality is a consequence of our emotional needs and responses as well as of our ability to rationally evaluate alternatives is strong enough to warrant making room for a more integrated perspective of morality that acknowledges its biological basis and emotional component as well as the role of cognition. (Fleck & de Waal, 2002, p. 21)

Despite the gathering data, there are still some who are skeptical about the use of evidence from primatology to inform our understanding of human morality. Fleck & de Waal (2002) respond to these critics in the following way:

> A chimpanzee stroking and patting a victim of attack or sharing her food with a hungry companion shows attitudes that are hard to distinguish from those of a person taking a crying child in the arms, or doing volunteer work in a soup kitchen. To dismiss such evidence as a product of subjective interpretation by 'romantically inspired naturalists' (Williams, 1989, p. 190) or to classify all animal behavior as based on instinct and human behavior as proof of moral decency is misleading.
> (p. 23)

From moral and social psychology, we also have surmounting evidence pointing to a sentimentalism view of morality. The Social Intuitionist Model developed by Jonathan Haidt describes our moral decisions as being rarely directly

caused by our moral reasoning capacities. Yet, as Haidt stresses, his claim is a descriptive one. It is a claim about how moral judgments are *actually* made, not about how moral judgments *ought* to be made. Baron (1998) has demonstrated that people following their moral intuitions often bring about nonoptimal or even disastrous consequences in matter of public policy, public health, and the tort system. A correct understanding of the intuitive basis of moral judgment may therefore be useful in helping educators design programs (and environments) to improve the quality of moral judgment and behavior (Haidt, 2001, p. 345).

Haidt also expresses one of the worries that emerge from a sentimentalist account of morality, stressing the importance of understanding its sentimentalist nature. Even if social psychological analyses of morality are restricted to the realm of descriptive claims, still they are of relevance for philosophers. If we are naturally endowed with emotions and inclinations that influence our behavior, we need to have the best possible understanding of what they are and how they work; at least so as to be able to foster the ones that should be cultivated and inhibit the ones that lead to immoral behavior.

Fleck & de Waal (2002) underline this point when they state that while there is no denial that we are creatures of intellect, it is also clear that we are born with powerful inclinations and emotions that bias our thinking and behavior. Human morality, as they say, needs to take human nature into account by either fortifying certain natural tendencies or by countering other tendencies (Fleck & de Waal, 2002). Moreover, by seeking out discourse partners who are respected for their wisdom and openmindedness, and by talking about the evidence, justifications, and mitigating factors involved in a potential moral violation, people can help trigger a variety of conflicting intuitions in each other. If more conflicting intuitions are triggered, the final judgment is likely to be more nuanced and ultimately more reasonable (Haidt, 2001, p. 355).

Haidt (2001), Greene (2001, 2004), Prinz (2007), and Nichols (2002, 2004) have all used empirical findings to challenge moral rationalist views. Haidt nicely outlines their perspective in the following passage:

> Now we know (again) that most of cognition occurs automatically and outside of consciousness (Bargh & Chartrand, 1999) and that people cannot tell us how they really reached a judgment (Nisbett & Wilson, 1977). Now we know that the brain is a connectionist system that tunes up slowly but is then able to evaluate complex situations quickly (Bechtel & Abrahamsen, 1991). Now we know that emotions are not as irrational (Frank, 1988), that reasoning is not as reliable (Kahneman & Tversky, 1984), and that animals are not as amoral (de Waal, 1996) as we thought in the 1970's. The time may be right, therefore, to take another look at Hume's perverse thesis: that moral emotions and intuitions drive moral reasoning. (Haidt, 2001, p. 355)

In addition to pointing in the direction of a more sentimentalist political philosophy, the empirical sciences have been enlightening the rationale behind our endorsement of particular moral principles. Some political philosophers, like Michael Walzer and, more recently, David Miller, have already called our attention to the impossibility of arriving at one single system of moral rules. They argue that human morality is inherently pluralistic: we make use of a variety of distinct principles in our moral judgments, according to the respective sphere of life in which the judgment is being made.

Interestingly, evolutionary biology has been providing philosophers with scientific evidence that such is indeed the case. According to Krebs (2002), human morality evolved in such a way that we inherited flexible programs that organize sets of conditional strategies. These strategies are domain-specific in the sense that they regulate distinct types of social relation – hierarchical, egalitarian, and intimate. Hence the pluralists may have gotten it right: we have evolved to endorse distinct moral principles when placed in different contexts.

Greene (2008) makes a different argument based on his research and on evolutionary explanations of moral judgments. He argues that the reason why we endorse deontological moral judgments is evolutionary: we came to develop strong emotional responses to human behaviors that were conducive to our survival and existence as a species. Yet Greene claims that, due to a change in our *modus vivendi*, many of these responses are no longer adaptive from an evolutionary perspective. Therefore, he contends that the understanding of our moral psychology casts doubt on deontology as a school of normative moral thought.

From political psychology we have also discussed recent research that adds support for a sentimentalist morality. I refer here to the motivated social-cognitive approach discussed in the preceding section. Political psychologists have been able to show that many of the political beliefs that we hold about, for instance, the justice of our economic system, are motivated by psychological needs of personal, group, and/or system justification. In this sense, these beliefs are less rational and more influenced by our psychological needs than one could anticipate.

Another interesting implication of the empirical sciences is provided by moral psychology and behavioral economics. Based on their respective researches, philosophers have begun to question the reliability of our moral intuitions, arguing that better understanding its psychological underpinnings has important normative implications (Baron, 1994; Horowitz, 1998; Unger, 1996; Sinnott-Armstrong, 2006).

For instance, Sinnott-Armstrong (2006) contends that empirical psychology has important implications for moral epistemology, which includes the study of whether, when, and how moral beliefs can be justified. When beliefs are justified depends on when they are reliable or when believers have reasons to believe that they are reliable. In circumstances where beliefs are based on processes that are neither reliable nor justifiably believed to be reliable, they are not justified. Psychological research, including research into framing effects, can give us reason to doubt the reliability of certain kinds of beliefs in certain circumstances. Such empirical research can, then, show that certain moral beliefs are not justified. Moral intuitionists cannot simply dismiss empirical psychology as irrelevant to their enterprise. They need to find out whether the empirical presuppositions of their normative views are accurate. They cannot do that without learning more about psychology and especially about how our moral beliefs are actually formed (Sinnott-Armstrong, 2006).

In the face of this and other similar interpretations of recent evidence on moral intuitions, Joyce claims that there is now a "general worry that empirical discoveries about the genealogy of moral judgments may undermine their epistemic status and ultimately detract from their authoritative role in our practical deliberations", going on to add that "This is a possibility to be taken seriously and explored carefully" (Joyce, 2008, p. 392). Some of the most prominent contemporary defenders of utilitarianism have taken this possibility seriously in arguing for their preferred ethical approach. Both Peter Singer and Peter Unger have objected to the authoritative role

conceded to our moral intuitions, supporting what they call a *Liberationist* approach to ethics – which, these philosophers argue, will ultimately result in the full endorsement of utilitarianism.

Lastly, findings in the field of behavioral economics have revealed some interesting implications for social policy. For example, the policy proposal specified in Nudge claims to be a solution to the problem of paternalism. Yet it demands more philosophical discussion than one may at first glance suppose. Now that behavioral economists and psychologists have documented several human inconsistencies in human decision-making, we are left with the following alternatives: either well-defined preferences do not exist, or we have conflicting preferences. In the face of this reality, Thaler & Sunstein (2009) claim, it is now possible to determine a person's best interest and to improve her wellbeing through the use of *libertarian paternalism*. Trout (2005) makes a similar case. He argues that in particular circumstances we are able to detect that a person's considered judgments or long-term goals are not being pursued due to the interference of some cognitive bias. Thus we can now use these very same biases to guide individuals in the direction of their long-term interests.

Nonetheless, a person's best interest still demands normative judgments to be defined, even with the help of the results of behavioral economics. Especially in the face of these results, we can no longer rely on the principle of revealed preference to empirically unveil an individual's best interest. Sugden (2006) emphasizes precisely this deontological point when he argues that the new findings from behavioral economics do not justify paternalism. In spite of the fact that empirical results point in the direction of the existence of incoherent preferences, Sugden claims that any form of paternalism would still threaten an important form of autonomy: namely, the opportunity to act based on unconsidered preferences. Even if we are choosing something that under reflection we would not choose, preserving the liberty to do so is an important form of freedom that should not be prevented. Hence the need for more informed philosophical discussion of this and other policy proposals that have been emerging from the empirical sciences.

Reclaiming Moral Sentimentalism in Political Philosophy

Here I will discuss the first implication of the empirical evidence discussed in the preceding paper, namely, that morality is a matter of sentiments as much as it is a matter of reason. Firstly, I will make the case for moral sentimentalism based on the relevant evidence gathered so far by empirical scientists. Secondly, I will argue that contemporary political philosophers have not properly acknowledged the relevance of moral sentimentalism due to a misinterpretation of the theory. As a consequence, the current literature on theories of justice is characterized by the pervasive presence of rationalism. Thirdly, I will expose the problems that political philosophers have mistakenly attributed to a moral sentimentalist approach to justice. These problems have stood in the away of a proper acknowledgment of the relevance of the theory. Fourthly, I will discuss the solutions to these problems. Lastly, I will argue that recent empirical evidence pointing to a sentimentalist nature of our morality combined with the fact that the problems attributed to the theory are not apt are good enough reasons for a sentimentalist turn in political philosophy.

Introduction

Contemporary political philosophers have assigned a secondary role to moral sentiments in the development of their theories of distributive justice. Of the two enlightenments that occurred in the eighteenth century, the rationalist and the sentimentalist,[39] the majority of contemporary political philosophers have fully embraced the rationalist one. For instance, one can find in Rawls a definition of "enlightenment liberalism [as] a comprehensive liberal and often secular doctrine founded on reason" (Rawls, Political Liberalism, p. xxxviii). In this sense, enlightenment liberalism encompasses a doctrine capable of supporting political morality via a direct appeal to our rational faculty alone.

The claim expressed in the preceding paragraph should not be understood as stating that all contemporary political philosophers are guilty of a complete disregard

[39] Frazer, *The Enlightenment of Sympathy*, 2010.

for our moral sentiments. Quite the contrary, many political philosophers have included moral sentiments in their discussions about justice. Yet this inclusion usually occurs late in the process of the development of their principles. A paradigmatic illustration of the role that contemporary political philosophers have assigned to sentiments in their theories can be encountered in Rawls' groundbreaking work *A Theory of Justice*. There we can find the analysis of our moral sentiments only towards the end of the book, more precisely in chapter eight, where Rawls investigates more closely our sense of justice. What is important to stress here is that this investigation is explicitly aimed at tackling the problem of relative stability, and is undertaken only after Rawls has already developed his principles of justice and provided them with a solid Kantian footing. Hence, Rawls interprets our affective structure as playing either one of the two following roles in theories of justice: as proving to be fit, or to constitute an obstacle to the application of justice principles.

In spite of this widespread reliance on our rational capacities as the ultimate ground for human morality, several empirical scientists have been gathering evidence that points in a different direction. The current findings suggest that our moral rules are to a large extent emotionally grounded. More recently, experimental philosophers have joined the effort of better understanding our moral nature and have been adding to the already existent data more evidence that our moral rules are less Kantian than the rationalist crowd could have anticipated.

For instance, Haidt reports extensive empirical support for the hypothesis that our moral judgments are triggered by emotional reactions, and that we are easily *morally dumbfounded* by our own moral intuitions (Haidt *et al*., 1993). A number of other experiments show that our moral judgments are strongly affected by features such as environmental clues, heuristics and biases, emotional intuitions, and the like (e.g. Cushman *et al*., 2006; Greene *et al*., 2004; Sinnott-Armstrong *et al*., 2010; Wheatley & Haidt, 2005). It is at the very least surprising that numerous contemporary political philosophers have remained alienated from moral sentimentalism in the face of accumulating evidence[40] pointing towards an emotional account of morality.[41]

[40] Vide second paper in this dissertation.

[41] Once again, it is important to clarify that political philosophers have not treated our affective states as completely irrelevant to the subject matter of justice; it is instead a claim that they have only acknowledged them insofar as they constitute an important step in the judgment of the stability of institutional arrangements – and this acknowledgment, as I will show, is insufficient.

Frazer (2010) argues that the dismissal of moral sentimentalism (and the embracement of moral rationalism) by political philosophers is a direct consequence of the fear of falling into a merely descriptive account of morality. Moreover, as I have argued in a separate paper, this fear has also kept contemporary political philosophers from assigning a more substantive role to empirical evidence about human moral behavior in the development of their theories. If we add these consequences together, i.e., the secondary role assigned both to moral sentiments and to empirical evidence in theories of justice, we naturally end up with a *rationalistic trend* in political philosophy that tends to perpetuate itself. This self-perpetuating tendency prevents rationalist philosophers from having to constantly keep up with the advancements of the empirical sciences.

Political philosophers have historically presented other reasons for the dismissal of moral sentimentalism. These reasons are: (i) *The Natural Fallacy Problem*; (ii) *The Stability Problem*; (iii) *The Problem of the Separateness of Persons*, and (iv) *Hume's Conservatism Problem*. In order to argue in favor of moral sentimentalism, I will address each of these problems in the following pages, showing that they should not stand in the way of a sentimentalist shift in political philosophy.

Some efforts to incorporate affect and emotion into the political philosophical debate have already been recently appearing in the literature. In this context, Michael Frazer, in his recent book *The Enlightenment of Sympathy*, attempts to begin the hard work of building a more sentimentalist view of justice. In his words,

> I seek to reclaim the sentimentalist account of reflection as a resource for enriching political science, political philosophy, and political practice today, a resource often overlooked due to the widespread influence of the opposed rationalist account.
> (Frazer, 2010, p. 4)

Along similar lines, the aim of this paper is to argue that a serious consideration of recent empirical evidence about human morality will lead to the fruitful embracement of moral sentimentalism by contemporary political philosophers. This emotional turn has already made a considerable impact in moral philosophy, but has yet to influence its political counterpart.

In order to reclaim a more substantive role for moral sentimentalism in contemporary political philosophy, I will firstly discuss the recent empirical findings

that taken together make a strong case in favor of a sentimentalist account of human morality. Secondly, I will highlight the absence of moral sentiments in contemporary theories of justice, arguing that this current state of affairs is a consequence of the rationalist enlightenment. Thirdly, I will undercut the arguments political philosophers have used to dismiss moral sentimentalist as a sound moral theory. Fourthly, I will discuss the solutions to the aforementioned problems. At last, and in light of all the previous discussions, I will wrap up the case for embracing moral sentimentalism in political philosophy, and I will discuss in a rather incipient manner some possible implications of such embracement.

2. The Empirical Case for Moral Sentimentalism

The case for moral sentimentalism has already been made by contemporary moral philosophers (Prinz, 2006; Nichols, 2004). As is the case with most philosophical quandaries, there is to date no consensus on the issue. Yet the recent sentimentalist revolution has already made a remarkable and undeniable impact on moral philosophy. Nonetheless political philosophy has remained largely unaffected by this emotional turn in our understanding of human moral systems. Therein lies the motivation for this section: the need to remake the empirical case for moral sentimentalism now in the realm of political philosophical theories.

For the fulfillment of this purpose, I will depict the relevant data following the argumentative structure developed by Jesse Prinz in *The Emotional Basis of Moral Judgments*. Firstly, I will discuss the evidence that points to the coexistence of emotions along with moral judgments – a hardly controversial claim. Secondly, I will take one step further in talking about the evidence that shows that our emotional states interfere with the moral judgments that we make. Thirdly, I'll move to a more controversial claim and discuss evidence that indicate that emotions can be a sufficient cause for moral judgment. At last, as a final case, I will argue that emotions are necessary for acquiring the capacity to make moral judgments and that emotions are also synchronically necessary for moral judgment. Despite the fact that this last case is weaker than the former ones, taken together these findings are sufficient to make a strong case for the view that emotions are more relevant to morality than currently assumed by contemporary political philosophers.

In regard to the first point, there are several studies showing that emotions are present when we make moral decisions. For instance, Sanfey, Rilling, Aronson, Nystrom & Cohen (2003) provide an analysis of brain imaging studies involving ultimatum games that reveal which areas of the brain are more active when individuals engage in distinct strategies. In ultimatum games, the proposer formulates a proposal for sharing money that can be seen either as fair or unfair by the recipient who, in turn, responds by accepting or rejecting it. Sanfey, Rilling, Aronson, Nystrom & Cohen (2003) report findings that show that unfair offers produce increased activity in the anterior insula, an area associated with the emotions of anger, disgust, and autonomic arousal.

Along the same lines, Joshua Greene gathers evidence from neuroimaging involving a series of moral dilemmas in which the participants have two morally unsound options to choose from, one being representative of consequentialist and the other of deontological morality. The dilemmas are variations of the trolley problem, including the footbridge and the loop case versions.[42] Greene reports that whenever participants opt for the deontological option, the brain regions activated are the ones responsible for emotions; and whenever the consequentialist alternative is chosen, the brain regions activated are the ones related with rational cognition. Greene (2008) interprets these findings as revealing the psychological rationale that underlies the main moral philosophical theories.

Additionally, Moll *et al.* (2001) and Moll *et al.* (2002) have provided cumulative evidence that the emotional parts of the brain are the more activated ones when people think about sentences with moral content. Moreover, Damasio et al. (1990) and Damasio (1994) have showed that brain damage can impair the development of cognitive-affective connections, consequently impairing the development of moral competence.

Now moving to the second point, I will discus the relevant data that show that our emotions are capable of interfering with our moral decisions. For example, Valdesolo & DeSteno investigate people's choices on trolley dilemmas and report that, in accordance with the thesis that negative affect is partly responsible for people's aversion to choosing the utilitarian alternative, participants who have just

[42] Foot (1978), Thomson (1976, 1985).

watched an amusing video clip are more likely to judge pushing the stranger off the bridge as more acceptable.

More direct evidence for this claim comes from studies where scenarios describing moral transgressions have been systematically manipulated, or relevant emotions have been experimentally induced, and moral judgment has been subsequently measured. Such studies typically find that moral violations are perceived as more or less severe depending on the perceiver's current emotional state, with direct consequences for attributions of blame and punishment. For instance, Goldberg *et al*. (1999) showed that witnessing a clear act of wrongdoing (e.g., watching a video of a man beating up a helpless teenager) triggers moral anger, which in turn increases punitiveness in subsequent judgments of unrelated transgressions (performed by a different perpetrator).

Turning to the emotion of disgust, Schnall *et al.* (2008) showed that subtly induced extraneous feelings of disgust increase the severity of moral judgments. Exposure to a bad smell, watching a disgusting film, and working in a dirty room all led participants to subsequently rate moral violations as more wrong, as compared to a control condition. This was especially the case for individuals who are more sensitive to their own bodily reactions and gut feelings. Similarly, a study investigating the effects of taste perceptions on moral judgments showed that consuming a bitter (as opposed to a sweet) beverage led to harsher judgments of moral transgressions (Eskine, Kacinik, & Prinz, 2011).

In this last study, researchers have also measured political attitudes and found that the reported effects emerged for political conservatives, but not for liberals. These findings are consistent with other works showing that conservatives are generally more sensitive to disgust (Haidt & Hersh, 2001; Inbar, Pizarro, & Bloom, 2009; Inbar, Pizarro, & Haidt, 2012). Conversely, Inbar et al. found that exposure to a disgusting odor led to more negative judgments of homosexuals (especially gay men) by both liberals and conservatives. Thus, there is evidence that physical disgust elicits moral disgust, thereby amplifying moral judgment. However this effect is, at least in some cases, moderated by sensitivity to bodily reactions and political ideology (Inbar, Pizarro, & Bloom, 2012).

On a distinct vein, but still related to how our moral intuitions can be altered by nonrational processes, there is the established research program on framing effects. Under this program, Kahneman & Tversky (1979) have provided an important

example of such framing effects in an experiment where the participants were faced with two equivalent disease-fighting programs, and they had to choose which was morally preferable. Both programs yielded the exact same expected results. Yet they were presented to participants under different frames, all in terms of probabilities: one program emphasized the number of people who would likely be saved, while the other program emphasized the number of those who would likely die. Instead of realizing that the results were the same and being indifferent to which program would be implemented, people always chose the one that expressed the least number of expected deaths (and the higher number of expected lives saved). As another example of this line of research, Unger (1996) shows that our intuitions about morality are subject to order effects. That is, our moral judgment shifts if two alternatives are presented either as a pair or as part of a list with additional alternatives.

Sinnott-Armstrong (2005) argues that the results from the aforementioned research on the nature of human moral beliefs and intuitions undermine moral intuitionism. He claims that the immediate implication of these results for moral philosophy is that all of our spontaneous moral beliefs cannot be justified non-inferentially and have to pass through a confirmatory process before we take them as credible.

An alternative interpretation of these results point in a rather different direction. Gill & Nichols (2008) take this evidence about the influence of emotions on our moral judgments to indicate that moral sentiments are at the basis of our commonsense moral judgments and that moral rationalists are now the ones carrying the burden of proof if they are to insist otherwise. They stress that

> (…) if we are right about the role of emotions in commonsense morality, then rationalists face a dilemma: either give up the claim that reason alone is the only proper ultimate ground of moral judgment, or give up the bulk of commonsense morality. (p. 153)

Moving to the third and more controversial point, I will now explore the data showing that emotions can be a sufficient cause for a moral judgment. The first relevant study to make this case is provided by Wheatley & Haidt (2005), who showed that moral judgments could be caused through the arousal of the feeling of disgust. In their experiment, whenever a story contained a word that elicited disgust participants were more likely to strongly morally condemn the acts under evaluation.

Here is how their study was implemented: they hypnotized participants to feel 'a brief pang of disgust (…) a sickening feeling in your stomach' at encountering an arbitrary word and then presented them with vignettes describing different moral offenses that either contained the target (disgust-related) word or not. The authors found that feelings of disgust (elicited by encountering the target word in the vignette) led to more severe moral judgments of the protagonist's actions (e.g., shoplifting, theft, bribery, incest). Most relevant to the moralization hypothesis, these effects were obtained even for a scenario that did not describe a moral violation (a scenario in which a student-council representative selected topics that would stimulate discussion for the upcoming meetings). Thus, subtly induced disgust influenced subsequent unrelated judgments and even moralized non-offensive acts.

The second line of research that suggests that emotions play a causal role in our moral decisions is the literature on *moral dumbfounding*. Haidt, Koller & Dias (1993), Haidt (2001), and Haidt & Hersh (2001) provide cross-cultural evidence that when people are confronted with harmless but supposedly offensive actions, or when they are questioned about issues related to sexual morality, they usually end up morally dumbfounded. That is, they "stutter, laugh, and express surprise at their inability to find supporting reasons, yet they would not change their initial judgment of condemnation" (Haidt, 2001, p. 346). In their experiments, the researchers ask individuals to judge the wrongness of a variety of moral dilemmas. After they have decided which acts are morally right or wrong, the individuals are given a series of arguments that invalidate all of the tentative reasons they try to provide in support of their moral judgments. The interesting result is that they end up admitting that they do not have *any* reason to hold that particular judgment, and yet they do not change their minds – they just know it is wrong, without knowing why.

Haidt (2007) argues that all these aforementioned empirical findings reveal four reasons for questioning the rationalist model of morality: (i) moral judgments involve the use of both reasoning and intuitive cognitive processes, and the former has been overemphasized; (ii) human reasoning appears to be always motivated; (iii) our experience of objective reasoning is frequently illusory – justifications are usually constructed by *post hoc* reasoning; and (iv) there is a higher degree of covariance between moral action and moral emotion than between moral action and moral reasoning.

Now to the final point, I will argue that emotions are necessary for acquiring the capacity to make moral judgments and that emotions are also synchronically necessary for moral judgment. So first I will discuss the data showing that emotions are diachronically necessary. The most decisive data so far are from the experiments with psychopaths on the distinction between moral and conventional rules. James Blair (1995, 1997) presents remarkable results from his developmental psychology research on the moral/ conventional distinction. He shows that, for the average person, moral rules such as 'do not hit other children' are understood to be wrong independently of any existent system of rules, and the given justification for this independence relies on arguments such as 'it hurts!' He also shows that, for the average person, conventional rules such as 'do not talk during class' are understood to be wrong only insofar as they are prohibited by the existent system of rules. Hence, normal children and adults perform well in tasks that involve the distinction between moral and conventional rules.

Conversely, Blair shows that both psychopaths and children with psychopathic tendencies perform atypically on these tasks. They are unable to draw the distinction between these rules, arguing that moral violations are equivalent to law violations – there is nothing morally wrong independently of the system of prevailing rules. The remarkable feature of this research is that psychopaths score perfectly normally on standard cognitive and intellectual measures, showing a diminishing capacity exclusively in terms of emotional responses to the suffering of others. These results show that emotional responsiveness plays a crucial causal role in the generation of normal moral judgments.

In addition to the research on psychopaths, another area of empirical inquiry has revealed a causal role for emotions in moral judgments. Researchers about trolley cases and other moral dilemmas have revealed a persistent finding: lay people draw a moral distinction that matches the philosophical view that it is permissible to divert the trolley in the bystander case but impermissible to push the man in the footbridge case. Interestingly, judgments made by patients with damage to the ventro-medial pre-frontal cortex (brain region associated with emotional sensitivity) to these same dilemmas consistently differ from those of the normal population (Koenig, Young, *et al*., 2007). The patients make no distinction among the cases, judging that it is equally permissible both to divert the train and to push the man. In this context, Gill & Nichols (2008) emphasize that "the evidence on psychopaths and patients with

ventro-medial damage suggests that emotions play a critical role in normal moral judgment – that without certain emotional responses, a person's moral judgment will be abnormal or incongruous" (p. 144).

At last, I turn my attention to the arguments in favor of attributing to emotions a synchronically causal role in the generation of moral judgments. These arguments are developed in Prinz (2006) and involve a *dispositional thesis* and an *anthropological argument*. About the former:

> Here, some caution is needed. Obviously, we can say things like, 'killing is wrong' without feeling any emotion. We have committed these rules to memory. It's a bit like reporting that bananas are yellow without forming a mental image of yellowness. The necessity thesis I have in mind is dispositional. Can one sincerely attest that killing is morally wrong without being disposed to have negative emotions towards killing? My intuition here is that such a person would be confused or insincere. (Prinz, 2006, p. 32)

Prinz asks us to imagine a person who is fully aware of all the non-emotional features of the act of killing. This person is also aware of the fact that killing decreases utility and that if it were to be universalized as a maxim such as 'thou shalt kill' it would lead to practical irrationality. Then Prinz asks us: would we think that this person holds the belief that killing is wrong? And his answer is that we would not think such to be the case. He claims that this person could have all the accurate beliefs about the objective features of the act of killing without having any clue about the meaning of the moral concept of wrongness. To wrap up this argument, Prinz adds:

> Conversely, if a person did harbor a strong negative sentiment towards killing, we would say that she believes killing to be morally wrong, even if she did not have any explicit belief about whether killing diminished utility or led to contradictions in the will. These intuitions suggest that emotions are both necessary and sufficient for moral judgment. (2006, p. 32)

About the anthropological argument, Prinz claims that, "if moral judgments were based on something other than emotions—something like reason or observation—we would expect more moral convergence cross-culturally. Reason and observation lead to convergence over time" (2006, p. 33). Yet what we find is pervasive divergence in moral beliefs across different cultures. To prove this point,

Prinz relies on an extensive review of cross-cultural moral divergence provided in one of his latest books, such as the ones described in the following passage:

> The Guhuku-Gama of New Guinea and other headhunters think it is okay to kill innocent people; the Greek citizens of Ptolemaic Egypt married their siblings at a rate of up to 30%; the Aztecs of Mexico and countless small-scale societies indulged in cannibalism; the Romans filled arenas to watch gladiators slaughter each other; Thonga men have sex with their daughters before hunting; the women of China endured excruciating pain by binding their feet; gender inequity and slavery have been widely accepted, and widely condemned. (2006, p.33)

Prinz recognizes that this last point is merely suggestive that moral values do not have an entirely cognitive foundation. Yet even if the widespread existence of moral divergence is not sufficient to directly demonstrate the necessity of emotions as a component of morality, it still serves as indirect evidence of this necessity. In other words, emotions do not only reinforce our antecedent sentiments or beliefs that a behavior is morally right or wrong, thus polarizing judgment, but they in addition may determine whether we identify the behavior in terms of morally right or wrong in the first place.

On the basis of these and other findings, Prinz (2006) argues that emotions can directly cause moral evaluations and that, unlike conventional rules, moral rules are fundamentally grounded in emotions. On his 'sentimentalist' view, believing that something is morally wrong is in essence having ''a sentiment of disapprobation'' towards it (Prinz, 2006, p. 33). In other words, condemning an act as immoral entails the experience of a negative emotional reaction, and the judgment itself is just an expression of this emotional reaction. Prinz contends that ''the emotion serves as the vehicle of the concept 'wrong' in much the same way that an image of some specific hue might serve as the vehicle for the thought that cherries are red'' (2006, p.34). Thus, Prinz's view can be seen as stating a Humean perspective of morality. In this sense, emotions do not only partly constitute our moral judgments, but they are also necessary for them.

Similarly, Nichols (2002) put forward a 'norms with feelings' account of morality, which holds that moral judgment is contingent on the interaction of two mechanisms: a system of rules (normative theory) prohibiting certain actions, and an independent affective mechanism that is activated by witnessing suffering in others.

In support of his 'affect-backed' theory, Nichols showed that certain non-harm-based transgressions that elicit disgust (like spitting in one's glass at a dinner party) are treated as nonconventional (i.e., moral) violations. That is, they become moralized: they were rated as less permissible, more serious, and more authority-independent than conventional offenses. In addition, it was demonstrated that the effects on the last two measures, seriousness and authority-independence, were stronger for individuals with high disgust sensitivity.

Although the exact mechanism driving the effects of emotion on moral judgment remain ambiguous, Nichols' account seems to imply, in line with the view defended by Prinz, that emotions are necessary for moral judgment. Moreover, it seems to imply that the intensity of one's affective reactions to a violation is crucial for whether the violation will be treated as a moral or a conventional one.

In the next section, I provide a (very) brief account of contemporary political philosophy, describing two of its main theories – namely, liberal egalitarianism and libertarianism. My goal is to illustrate how its development has been segregated from moral sentimentalism. As an example of this segregation, I discuss the role played by emotions in Rawls's theory of justice as the paradigmatic case of which role emotions have been relegated to in the theories originated in the realm of the enlightenment of reason.

3. The Emergence of Reason and the Annihilation of Sentiments: the historical grounds

Much of contemporary political philosophy draws on the seminal work of John Rawls. Be it to reverence or to criticize it, every political philosopher is inevitably compelled to address the arguments presented in his book *A Theory of Justice*. In the same vein as the Rawlsian theory, political philosophers in general have welcomed and embraced the Kantian rationalist approach to understand matters of justice. Since the publication of Rawls's groundbreaking work four decades ago an immense variety of deviations and improvements of his original justice principles have been proposed and discussed in the political philosophical literature. However, it remains unclear whether anything parallel to a consensus will ever be arrived at within the boundaries of rationalism.

Rawls sustains that one of the main tasks of political philosophers is to build a coherent system of social cooperation capable of accounting for the coexistence of the most diverse worldviews. Thus, it follows that political philosophy ought to involve the devising of a feasible form of political liberalism, so as to peacefully and productively accommodate pluralism in terms of moral, religious, cultural and philosophical beliefs. In this context, a crucial part of any political system of cooperation is the establishment of principles of distributive justice; principles concerned with the manner in which economic benefits and burdens are distributed across individuals in society. The goal of distributive justice is to provide the necessary moral guidance in developing political processes and structures that impact the distribution of economic outcomes.

As already pointed out, contemporary political philosophers have developed several theories of distributive justice. These theories vary across the many dimensions that comprise distributive principles, such as: (i) what is relevant – income, wealth, opportunities, jobs, welfare, utility, etc.; (ii) the nature of the recipients – individuals, groups, classes, etc.; and (iii) on what basis should the distribution be made – equality, maximization, according to individual characteristics, according to free transactions, etc. Nonetheless the point is that the majority of these theories remain constant along one fundamental dimension: rationalism.

The main contemporary theories of distributive justice can be divided under two broad categories: (i) *liberal egalitarianism*, and (ii) *libertarianism*. On the one hand, libertarians are solely concerned with the protection of individual rights, firstly envisaged by Locke as the natural rights to life, liberty and property. On the other hand, liberal egalitarians include all political philosophers whose theories also share the libertarian embracement of the intrinsic value of autonomy and the consequent relevance of individual liberties, while at the same time acknowledging the injustices engendered by the existent discrepancies in human social and economic conditions.

Under this division, the main liberal egalitarian models of distributive justice are defined by *strict egalitarianism*, the *difference principle*, and *luck egalitarianism*. Strict egalitarians advocate for the equal allocation of material goods to all members of society, grounded on the philosophical claim that people are morally equal and the best way to give effect to this moral ideal is the equal distribution of all material goods. This is a difficult position to defend, given that equal distributions are easily

criticized in the face of the economic fact that everyone can be made materially better off if incomes are not strictly equal amongst individuals (Carens, 1981).

The difference principle is one of the main features of Rawls's theory of justice, and spells out the way in which he interprets the fairness of a distribution of goods. As it stands, the difference principle states that it is acceptable to diverge from strict equality as long as the inequalities that follow improve the conditions of those least advantaged in society. This principle shares a similar ground with strict egalitarianism. That is, they are both grounded on the premises of equal respect for persons and the denial of moral desert. In order to clarify this similarity, it suffices to show that the difference principle collapses to a form of strict equality in a world where differences in income have no effect on the incentives necessary for people to work.

At last, luck egalitarians focus on the moral roles of luck and responsibility in the economic life. They intend to provide a critical response to Rawls's approach given the minor role to which Rawls relegates responsibility in his theory. The most prominent advocate of this view is Ronald Dworkin, who elaborated his theory based on the concept of equality of opportunity and on the following distinction amongst the different sources of social and economic benefits and burdens:

(i) Brute luck (endowments): not a matter of deliberate gambles, such as genetic inheritance, unforeseeable bad luck, etc.; and

(ii) Option luck (choice, ambitions): a matter of how deliberate and calculated gambles turn out – whether someone gains or loses through accepting an isolated risk he should have anticipated and might have declined, such as the choice to work hard, to spend money on expensive luxuries, etc.

Based on this distinction, Dworkin is capable of attributing a larger role for individual responsibility in the distribution of social and economic outputs. According to his view, individuals are shielded from the results of brute luck, but are considered responsible for how things turn out for them as a consequence of option luck.

I have briefly described the main contemporary theories of distributive justice so as to illustrate to which extent moral sentiments are currently absent from the political debate. My hypothesis here is that this present state of affairs is the result of

a historical episode.[43] Despite the fact that in the eighteenth century both rationalist and sentimentalist accounts of autonomous reflection had many worthy advocates, the majority of contemporary political philosophers hold a commitment to individual autonomy most often understood in Kantian, rationalist terms. That is, individual autonomy is identified with the individual exercise of reason. The most prominent political philosopher of our time was not immune to this attitude; when he wrote his masterpiece *A Theory of Justice* Rawls explicitly presented his project as a Kantian one.

In this context, Michael Frazer (2010) insightfully elucidates the rationale behind this widespread rationalist attitude among political philosophers. He argues that the study of eighteenth century moral and political thought reveals that there were in fact two coexistent Enlightenments concerning the analysis of moral and political reflection. The first is the one he calls the *Rationalist Enlightenment*, corresponding to the common conception of the eighteenth century as the age of reason. The second is the one he calls the *Sentimentalist Enlightenment*, corresponding to an age not only of reason but also of "reflectively refined feelings shared among individuals via the all-important faculty of sympathy" (p. 4). However, Frazer (2010) stresses an important caveat in this diagnosis of eighteenth century philosophy:

> This is not to say that every moral and political thinker of the Enlightenment can be easily classified as exclusively 'rationalist' or 'sentimentalist'. Many of the greatest thinkers of the period – most notably Jean-Jacques Rousseau – evade such simple categorization. But there was clearly an ongoing debate in the eighteenth century over the nature of reflective autonomy – a debate in which many took an identifiably rationalist position and many others an identifiably sentimentalist one. (p. 4)

[43] In line with the argument developed by Frazer (2010).

In addition, Frazer (2010) remarks:

> Most of the major philosophers of the sentimentalist Enlightenment – such as the Third Earl of Shaftesbury, Joseph Butler, Francis Hutcheson, David Hume, and Adam Smith – were British, while many of the major rationalists of the period were French or German. It is important, however, not to confuse the distinction between the rationalist and sentimentalist Enlightenments with the distinctions that have been drawn among the various 'national' Enlightenments. There were many rationalists in Britain – among them Samuel Clarke, William Wollaston, and Richard Price. There were also many sentimentalists on the continent, most notably J.G. Herder and, at least for a time, his teacher the precritical Immanuel Kant. (p. 3)

Hence it is important to comprehend sentimentalism and rationalism not as national worldviews, but as rival positions on a transnational debate. Most importantly, the two theories were part of a debate about the nature of reflective autonomy. A debate that was central in the intellectual life of the eighteenth century, and that remains central in political philosophy this day.

In order to clarify the sentimentalist as well as the rationalist Enlightenment, it is helpful to understand their competing theories of moral and political reflection as combining two separate elements. To use Hume's most famous distinction, they offer both a theory of what *is* and a theory of what *ought* to be. That is, they offer both a descriptive moral psychology and a theory of normativity. Regarding our descriptive moral psychology, sentimentalist philosophers describe our process of moral reflection as a matter of feeling and imagination as well as a matter of cognition. In contrast, rationalist philosophers describe human moral reflection solely as a matter of rational cognition. Concerning the theory of normativity, sentimentalists claim that normative force stems from the reflective stability of a mind able to bear its own holistic survey; while rationalists argue that the normative power of moral rules follow from the authoritative legislation of the human faculty of reason.

Rationalists from Plato onward have maintained that reason is rightly the master and passion is rightly the slave (Frazer, 2010). In this respect, Hume famously counter argued that 'reason is, and ought only to be, the slave of the passions'. Yet this remarkable passage, when taken in isolation, leads to a distorted understanding of Hume's actual view. Even though philosophers may be right in drawing a distinction

between reason and passion, Hume consistently maintains that the two are in reality 'uncompounded and inseparable'. This is not to deny Hume's point that reason alone is incapable of motivating action; this is a claim that the sentiments that Hume describes as being action-motivating are not to be comprehended merely as passions. When Hume refers to sentiments he refers to products of the mind as a whole, reason and imagination included. In this context, Frazer (2010) argues that

> (…) the contrast between rationalism and sentimentalism is best understood as the contrast between a hierarchical view of the moral soul, on one hand, and an egalitarian view, on the other – an egalitarian view in which normatively authoritative standards are the product of an entire mind in harmony with itself. (Frazer, 2010)

The enlightenment-era rationalists do not cast out human passions from their descriptive psychology. Instead rationalists argue for a secondary role for the passions in our psychic regime; they are to abide by the sovereign faculty of reason. In this sense, the duties of their station involve keeping quiet during the purely rational process of proper moral and political reflection, then deferring to the rationally authoritative principles that result from this process.

Just as the passions take a subordinate place in the rationalist psychic regime, the study of these non-rational forces takes a subordinate place in rationalist moral and political theory. For rationalists, empirical anthropology is always subsidiary to the a priori metaphysics of morals. That is, only after reason has finished determining what standards we ought to follow can we then address the empirical question of how social and psychological contingencies may be better brought in line with reason's authoritative demands.

In this sense, a rationalist approach to morality renders the knowledge from the empirical sciences almost unnecessary for political philosophical theorizing. This relative independence from the empirical sciences is precisely that which grants rationalism a worrisome self-perpetuating character. The rationalist approach frees the philosopher from having to constantly keep up with empirical evidence, resulting in an increasing distance between the empirical sciences and theories of justice. That is, the supremacy of reason is closely related to the diminished relevance of empirical work and, in turn, this diminished relevance is closely related to the perpetuation of the supremacy of reason.

In contrast with rationalism, a sentimentalist approach to morality assigns empirical evidence a crucial part in the development of moral theories. That is, sentimentalist philosophers begin their work where rationalists end it, namely, with the empirical examination of what actually motivates us to follow our current standards and practices.

For instance, rather than being presented as a possible normative justification of his theory, the discussion of moral development by Rawls in chapter 8 of *A Theory of Justice* is meant merely to counter one possible objection to justice as fairness: namely, that it might prove unstable over time. Concerning this point, Rawls (1971) writes:

> One conception of justice is more stable than another if the sense of justice that it tends to generate is stronger and more likely to override disruptive inclinations. In order to achieve such stability, it is thus critical that when institutions are just [as defined by this conception], those taking part in these arrangements acquire the corresponding sense of justice and desire to do their part in maintaining them. (…) However attractive a conception of justice might be on other grounds, it is seriously defective if the principles of moral psychology are such that it fails to engender in human beings the requisite desire to act upon it. (p. 398)

Hence Rawls does not intend the description of our moral psychology as part of the reflective justification for his conception of justice. Rawls (1971) goes on to add that "the main grounds for the principles of justice have already been presented. (…) At this point we are simply checking whether the conception already adopted is a feasible one and not so unstable that some other choice might be better" (p. 441). This attribution to our empirical psychology of a subsidiary place in moral and political theory is, as already discussed, one of the principal characteristics of the rationalist enlightenment – one that remains present and alive to this day.

In the next section I discuss the reasons that lead contemporary political philosophers to embrace the rationalist enlightenment, arguing that philosophers took the rationalist road due to a misinterpretation of moral sentimentalism. In order to clarify this misinterpretation, I will address all the problems that were mistakenly identified with a sentimentalist approach to morality.

4. The Alleged Problems with Moral Sentimentalism

Up to this point I have argued that the main reason why political philosophers have been struggling to incorporate more empirical research in to the development of their theories of justice is at least in part historical, going back to the eighteenth century enlightenments. From the two main movements of that century, the rationalist and the sentimentalist, contemporary political philosophy seems to have almost fully embraced the former. Yet the rationale behind this choice remains to be explained. What is it about moral sentimentalism that has lead to its disfavor among philosophers? And, most importantly, were they justified in refusing moral sentiments a more prominent role in the political sphere? I will elaborate on the former question in this section, and on the latter in the following section.

Philosophers have identified four main problems with moral sentimentalism: (i) *The Natural Fallacy Problem*; (ii) *The Stability Problem*; (iii) *The Problem of the Separateness of Persons*, and (iv) *Hume's Conservatism Problem*. In what follows, I will elucidate each of these worries.

(i) *The Natural Fallacy Problem*

This problem was first exposed by Moore in his *Principia Ethica* (1903) and can be spelled out in its simplest form as the impossibility of deriving *ought* statements from *is* statements. The argument consists in the thesis that no account of the development of our moral psychology could ever, by itself, justify our moral commitments; to believe otherwise is to confuse an empirical explanation of the origins of a value commitment with a demonstration of its genuine normative authority. Moore (1903) claims to have proved that positive moral claims do not follow from descriptive premises, along with the immediate implication that all the empirical sciences are irrelevant to moral philosophical theorizing and common moral beliefs. What we ought to do and how we decide this is a separate question from why and how moral systems arose.

Regarding the natural fallacy and the related threat to normative authority posed by moral sentimentalism, Korsgaard writes in *The Sources of Normativity*, "to raise the normative question is to ask whether our more unreflective moral beliefs and motives can withstand the test of reflection" (1996, p. 47). The fear derived from theories of the natural origin of our moral systems is, she argues, the reason why "we

seek a philosophical foundation for ethics in the first place: because we are afraid that the true explanation of why we have moral beliefs and motives might not be one that sustains them" (1996, p. 49). In these passages Korsgaard reveals a common view among moral and political philosophers, the view under which moral sentiments are understood as lacking any normative force and, therefore, as incapable of providing a proper ground for morality.

(ii) *The Stability Problem*

The stability problem is related to the fact that our sentiments, be they moral or not, are not characterized by immutability. That is, our sentiments are subject to biases, such that the same person can display dissimilar affective reactions when facing similar situations. This is a problem within the same individual; and distinct persons can display different affective reactions when facing the same situation – this is a problem across individuals.

The instability of our moral sentiments poses a problem to philosophers who are concerned with the necessity of universal moral principles. One of the most debated topics in the 'sentimentalist versus deontological morality' literature is the supposed necessity emphasized by Kant of having principles that are valid for all rational creatures. Deontology comprehends universality as one of the main features from which our moral rules derive their normative force. Yet despite all efforts undertaken by philosophers to provide such universal principles, their attempt has repeatedly failed in the face of a reality where moral systems portray widespread cultural variation.[44]

Hume, in *A Treatise of Human Nature*, has already called our attention to the first important bias to which our moral sentiments are susceptible. This bias is characterized by a propensity to be more tolerant towards those close and dear to us, and more hardhearted towards those who bare no connection with us. This is a direct consequence of the fact that our moral sentiments are derived via the mechanism of sympathy and that we naturally have stronger sympathy for those we cherish.

This human bias towards family and social acquaintances inevitably leads to interpersonal variation in the moral assessment of a person's action. This variation poses a problem: how to settle the disagreement about the moral status of an action

[44] Jesse Prinz, *The Emotional Construction of Morals*, 2007.

(or an individual) when the persons judging it are themselves unable to reach a consensus? In this context, Frazer (2010) points that:

> If moral evaluation is purely a matter of sentiment, however, it is unclear how this disagreement is to be adjudicated. Sentimentalism, according to a certain clichéd line of thought, leaves us no more able to account for our moral evaluations than we can for our aesthetic tastes, since it places moral virtue, like physical beauty, in the eye of the beholder. (p. 45)

This 'clichéd line of thought' mentioned by Frazer is a very common stance in moral and political philosophy. Sentiments are understood as mere emotional reactions to external situations, devoid of any cognitive features that would render them the subject of reasonable debate. In this sense, they would be comparable to our aesthetic tastes, therefore left to the sphere of private individual choices that are not subject to rational scrutiny. Nonetheless this view represents both a poor understanding of our aesthetic tastes and an inaccurate account of our moral sentiments.

Hume acknowledges that, once we recognize the pervasive divergence of our moral judgments, it is no more than natural to search for a shared 'standard of taste' (EMPL, p. 229). Yet he is also aware of the fact that this search is futile if we hold a relativistic perspective on human morality – one parallel to the one advocated for in the *Treatise*. This misguided line of reasoning would assert that:

> All sentiment is right; because sentiment has a reference to nothing beyond itself, and is always real, wherever a man is conscious of it. (…) One person may even perceive deformity, where another is sensible of beauty; and every individual ought to acquiesce in his own sentiment, without pretending to regulate those of others. (EMPL, p. 230)

The important thing to notice is that Hume's moral sentimentalism does not amount to a relativistic view of human morality. Quite the contrary, as we will discuss in the next section, he argues against moral relativism and it is clear that he is fully aware of the relevance of shared rules for achieving a well-functioning society.

Rationalists contest this interpretation of moral sentimentalism. Instead, they argue that all aspects of the human mind are contingent, with the sole exception of

pure reason – the sovereign faculty to which all others should be subservient. Frazer nicely captures their view in the following passage:

> Although the other features of the mind and personality are plagued by contingency, reason deals only with necessary truths. Although my emotions, imagination, and memory are all part of causal nexuses both natural and social, my reason is free. If I am to think of myself as free from natural and social contingency, Enlightenment rationalists maintained that I must think of my true self as purely rational. If my actions and my standards of action are to be truly my own, they maintained that it is this real self that must be sovereign, legislating standards in reflections and dictating behavior in practice. And since the true self is identified with a single faculty we are all held to share, the true self of all individuals is fundamentally the same. (Frazer, 2010, p. 6)

Sentimentalists nevertheless do not succumb to the rationalist's critique. Quite the contrary, they are immune to it. Sentimentalists endorse a distinct way of thinking about contingency. In opposition to rationalists, sentimentalists do not identify one's true self with a single human faculty. They identify one's true self with the entire set of human faculties: rational, social, and psychological aspects all included. As I will argue in the next section, this element of moral sentimentalism is going to be crucial in order to properly address the rationalists' claim that moral sentiments are unstable and, therefore, unworthy of being a ground for human morality.

Before we end the description of the stability problem, it is important to better spell out the rationalist's perspective on the issue. Unlike some of the more extreme rationalists, Kant and other rationalist philosophers rarely denied that social and psychological contingencies are responsible for much of our behavior. In this manner, they have not aimed at the extirpation of contingency form human life. Rather they have tried to bring all contingent forces under rational control, so that these forces guide us to the very same standards that reason necessarily and authoritatively demands. That is, the "Enlightenment-era rationalist position is generally Platonic, not Stoic; the passions are not to be banished from the psychic regime, but are to obey their superiors and keep to their proper place" (Frazer, 2010, p. 7).

(iii) *The Problem of the Separateness of Persons*

The problem of the separateness of persons was elaborated by Rawls in *A Theory of Justice* and is related to the contented inability of an ethics based on

sympathy (as exemplified by Hume and Smith) to properly acknowledge the inviolable dignity of each individual. The starting point of Rawls's argument is the conception of right that he takes to be shared by all moral sentimentalists, that is, that "something is right, a social system say, when an ideally rational and impartial spectator would approve of it from a general point of view should he possess all the relevant knowledge of the circumstances" (1971, p. 161).

Rawls claims that the above conception is hollow; nothing follows from it without further specification of the psychological features of the spectator. It is here that sympathy and the moral sentiments come into play, for Rawls notes that we might plausibly imagine the spectator to be a 'perfectly sympathetic being.' If we do so, Rawls believes we will find 'a natural derivation of the classical principle of utility' (1971, p. 162). The rationale behind this argument is as follows: if we can successfully imagine an omniscient entirely rational and perfectly impartial being, then we can also imagine that this being is endowed with sympathy so great as to feel the sentiments of all individuals with all of their original vehemence. Such an ideal spectator 'identifies with and experiences the desires of others as if these desires were his own' (Rawls. 1971, p. 24), psychologically fusing with the object of his sympathy. Let us now imagine that this spectator is asked to approve or disapprove of a given social system. He sympathizes fully with each person within this system and, as each in turn mingles with the spectator, they come to be united in a single psyche. As a result, Rawls concludes, it is such an impartial spectator 'who is conceived as carrying out the required organization of the desires of all persons into one coherent system of desire; it is by this construction that many persons are fused into one' (1971, p. 24). Rawls argues that this result is unacceptable insofar as it implicates the violation of one of our most cherished intuitions: the sanctity of the individual. This violation would also render the institutional system unstable, given that in many instances an individual might be required to sacrifice his welfare for the good of society.

The problem with Rawls's critique of moral sentimentalism is that he inadvertently lumps together the theories of David Hume and Adam Smith. In light of this error, Frazer (2010) argues that while Rawls is correct in addressing this critique to Hume, he is mistaken in extending it to Smith. As we will see in the next section, Adam Smith's theory is able to at the same time avoid this critique and maintain its commitment to moral sentimentalism.

(iv) *Hume's Conservatism Problem*

> John Stuart Mill wrote that Hume's 'absolute skepticism in speculation very naturally brought him round to Toryism in practice.' A philosophical skeptic in naturally a political conservative, Mill explains, because if 'one side of every question is about as likely as another to be true, a man will commonly be inclined to prefer that order of things which, being no more wrong than every other, he has hitherto found compatible with his private comforts.' (Frazer, 2010, p. 65)

Mill's depiction of Hume as a political conservative in virtue of his philosophical skepticism has been the subject of voluminous scholarly debate. Currently, there is no more agreement on the degree of Hume's conservatism than there is on the degree of his skepticism, let alone on the degree of connection between the two. Yet by now it should be clear that, at least regarding human morality, Hume was not the destructive skeptic depicted by Mill.

Philosophers have used the term 'justice' in a wide variety of ways. Sometimes justice is understood primarily as a virtue of social systems, at other times as a virtue of individuals. Sometimes justice becomes a sort of catchall term for the virtues generally – whether of individuals or societies – while at other times its usage is far narrower. Hume uses the term 'justice' in a very specific sense, i.e., to designate the individual's virtue of obedience to the rules that allow for social cooperation, particularly in the economic sphere. Hume's understanding of justice thus has the benefit of tying the character trait he identifies as the justice of individuals to features of the social systems under which an individual lives.

Yet Hume refrains from describing society's rules themselves as just or unjust – one of the most troubling features of his conception of the term. Social rules are often deemed cruel or useless, and Hume does not hesitate to call for their reform when such is the case. However the rules of justice are never unjust per se, at least not in Hume's use of the term. Also troubling is that the social rules at issue when Hume discusses justice are almost exclusively those governing the accumulation and exchange of property. In relation to today's use of the term, Hume's concept may seem far too constricted on this point.

In order to better understand the origin of Hume's political conservatism, we have to go back to his account of the origin of the artificial virtues. For Hume, artificial virtues such as justice involve obeying conventional rules, which Hume sees

as human creations, while natural virtues such as benevolence involve no such conventions. Yet it is important to stress that this interpretation does not amount to stating that justice is not natural, "Mankind is an inventive species, and where an invention is obvious and absolutely necessary, it may as properly be said to be natural as anything that proceeds immediately from original principles, without any thought or reflection" (Hume, T 3.2.2.19).

In the *Treatise* Hume emphasizes that "our sense of every kind of virtue is not natural; but (…) there are some virtues, that produce pleasure and approbation by means of an artifice or contrivance, which arises from the circumstances and necessities of mankind" (T 3.2.1.1). Hence Hume divides the virtues into those that are natural, in that our approval of them does not depend upon any cultural inventions or jointly-made social rules, and those that are artificial, dependent both for their existence as character traits and for their ethical merit on the presence of conventional rules for the common good. Following this division, he gives separate accounts of both kinds of virtues.

The traits he calls natural virtues are more refined and completed forms of those human sentiments we could expect to find even in people who belonged to no society but cooperated only within small familial groups. The traits he calls artificial virtues are the ones we need for successful *im*personal cooperation, for our natural sentiments are too partial to give rise to these without intervention. In this context, Hume declares that "the sense of justice and injustice is not deriv'd from nature, but arises artificially… from education, and human conventions" (T 3.2.1.17).

In the *Treatise*, Hume includes among the artificial virtues honesty with respect to property (which he often calls equity or "justice"), fidelity to promises (sometimes also listed under "justice"), allegiance to one's government, conformity to the laws of nations (for princes), chastity (refraining from non-marital sex) and modesty (both primarily for women and girls), and good manners. A great number of individual character traits are listed as natural virtues, but the main types discussed in detail are greatness of mind ("a hearty pride, or self-esteem, if well-concealed and well-founded," T 3.2.2.11), goodness or benevolence (an umbrella category covering generosity, gratitude, friendship, and more), and such natural abilities as prudence and wit, which, Hume argues, have a reasonably good claim to be included under the title moral virtue, though traditionally they are not.

Hume next poses two questions about the rules of ownership of property and the associated virtue of material honesty: what is the artifice by which human beings create them, and why do we attribute moral goodness and evil to the observance and neglect of these rules? The first half of the story starts with the fact that human beings naturally have many desires but are individually ill equipped with strength, natural weapons, or natural skills to satisfy them all. Yet humans can remedy these natural defects by means of social cooperation, i.e., the combination of strength, the division of labor, and the mutual aid in times of individual weakness. It occurs to people to form a society as a consequence of their experience with the small family groups into which they are born, groups united initially by sexual attraction and familial love, but in time demonstrating the many practical advantages of working together with others.

However, there is a problem. In the conditions of moderate scarcity in which humans find themselves, and in the face of the portable nature of the goods we desire, our untrammeled greed and our naturally "confined generosity" (generosity to those dear to us in preference to others) together tend to create conflict and undermine cooperation. This conflict may result in the destruction of collaborative arrangements among people who are not united by ties of affection and, as a consequence, leave us all materially poor. For Hume, no remedy for this natural partiality is to be found in "our natural uncultivated ideas of morality" (T 3.2.2.8). Therefore an invention is needed.

Hume argues that we originally create the rules of property ownership in order to satisfy our avidity for possessions for our loved ones and ourselves. Within small groups of cooperators, individuals signal to one another a willingness to conform to a simple rule: to refrain from the material goods others come to possess by labor or good fortune, provided those others will observe the same restraint toward them. This signaling is not a promise (which cannot occur in the absence of another, similar convention), but an expression of conditional intention. The usefulness of such a custom is so obvious that others will soon catch on and express a similar intention, and the rest will fall in line. The convention develops tacitly, such as the conventions of language and money. When an individual within such a small society violates this rule, the others are aware of it and exclude the offender from their cooperative activities. Once the convention is in place, justice is defined as conformity with the convention, and injustice as violation of it. The convention defines property rights,

ownership, financial obligation, theft, and related concepts, which had no meaning before the convention was "invented."

So useful and obvious is this invention that human beings would not live for long in isolated family groups or in fluctuating larger groups with unstable possession of goods. Their inventiveness would end up enabling them to come up with the rules of property, so as to be able to reap the economic benefits of cooperation in larger groups. Under the observance of these rules, people are capable of better satisfying their powerful natural greed via its regulation by the rules of justice.

Hume's interpretation of justice in this rather narrow sense and as an artificial versus a natural virtue is a consequence of the fact that his concept of sympathy is excessively simplistic. He conceives of sympathy merely as what we today call emotional contagion, and does not account for more sophisticated kinds of sympathetic emotions. Thus he misses the fact that justice can indeed be the direct product of sympathy, just not of the kind he works with – namely, emotional contagion.[45]

About the motivation to act in accordance to the rules of justice, Hume argues that we can have no natural motive, independent of al human invention, for governing our behavior according to such rigid rules. So in order to understand both the motivation we have for being just and the moral approbation we feel towards just individuals, Hume argues that we must not consider the individual acts demanded by justice in isolation, but rather as part of a larger pattern of behavior governed by artificial rules. This approach has been nicely elucidated by Rawls. On his view, practices are:

> (…) set up for various reasons, but one of them is that in many areas of conduct each person's deciding what to do on utilitarian grounds case by case leads to confusion, and that the attempt to coordinate behavior by trying to foresee how others will act is bound to fail. As an alternative one realizes that what is required is the establishment of a practice, the specification of a new form of activity; and from this one sees that a practice necessarily involves the abdication of full liberty to act on utilitarian and prudential grounds. It is the mark of a practice that being taught how to engage in it involves being instructed in the rules which define it, and that appeal is made to those rules to correct the behavior of those engaged in it. (…) Thus it is essential to the notion of a practice that the rules are publicly known and

[45] As will be discussed later, justice can be understood as a direct product of the Smithian conception of sympathy.

> understood as definitive; and it is essential also that the rules of a practice can be taught and can be acted upon to yield a coherent practice. On this conception, then, rules are not generalizations from the decisions of individuals applying the utilitarian principle directly and independently to recurrent particular cases. On the contrary, rules define a practice and are themselves the subject of the utilitarian principle. (1955, p. 24)

The sphere of life governed by the artificial virtues is in this respect like a sport or a game. It is no more possible to understand the motive for 'stealing a base' independent of the practice of baseball than it is to understand the motive for 'repaying a loan' independent of the practices of promising and property-exchange. The good that stems from an artificial virtue comes not from each individual virtuous act, but from the existence of the practices under whose rubric these acts occur.

Rather than describing a great legislator or a group of social-contractors deciding on which rules should govern human society, Hume presents the gradual development of these rules as an evolutionary byproduct of the correction of humanity's sense of self-interest. In this sense, the rules of justice provide the means to extend cooperation from the sphere of those who we strongly cherish to those who lay outside this sphere. We extend our cooperative behavior not out of a benevolent concern for those others, but because such cooperation is in our long-term self-interest. Likewise, it is in the long-term self-interest of those with whom we cooperate.

Hence the theory of justice we encounter in Hume's work does not specifically determine the content of the rules. As Frazer (2010) points out "A reflectively corrected sense of self-interest will necessarily lead us to embrace some system of justice and government, but it will not determine the specific shape this system takes." He goes on to point out that "Hume argues that we should almost always rest satisfied with whatever system under which we happen to find ourselves living" (Frazer, 2010, p. 71).

In this context, some philosophers have maintained that Hume is entirely indifferent to the particular form that the rules of justice take, as well as to the particular kind of government that enforces these rules. While this is clearly an exaggeration, the precise nature and degree of Hume's complacency with regard to existing political institutions is to this day the subject of active debate.

Hume argues that just actions are rarely immediately agreeable and that their utility is rarely obvious to the untrained eye. Given that individual acts of justice, taken in isolation, are often deleterious, it takes considerable reflection to shift our attention from the effects of individual actions to the effects of the larger, rule-governed practices of which they are a part. Hume describes this shift in the following passage:

> We partake of their uneasiness by sympathy, and as every thing, which gives uneasiness in human actions, upon the general survey, is called vice, and whatever produces satisfaction, in the same manner is denominated virtue, this is the reason why the sense of moral good and evil follows upon justice and injustice (…) Thus self-interest is the original motive to the establishment of justice: But a sympathy with the public interest is the source of the moral approbation, which attends that virtue. (Hume, T 3.2.2.24)

We now need to recall that our need for justice is derived from the natural biases in our sympathy, which lead us to promote our own good and the good of those close to us at the expense of those for whom we feel little or no concern. Likewise Hume claims that "if men had been endowed with such a strong regard for the public good, they would never have restrained themselves by these rules" (Hume, T 3.2.2.18). That is, were the limitations of our sympathy ever to be radically overcome, i.e., were we ever to become 'so replete with friendship and generosity that every man has the utmost tenderness for every man and feels no more concern for his own interest than for that of his fellows,' then 'the use of justice would, in this case, be suspended by such an extended benevolence' (EPM 3.1.6). Justice, in other words, is a virtue only among those who are not perfectly virtuous, and wins the approbation of our corrected, unbiased moral sentiments only insofar as we acknowledge the strongly biased nature of our moral sentiments.

On the one hand, in the liberal tradition, justice is identified with a commitment to interpersonal fairness. On the other hand, Hume goes in the opposite direction when he identifies justice with the strict adherence to the conventional rules of one's society. However, Hume's conservatism does not straightforwardly imply that one could not possibly achieve a roughly Rawlsian conception of justice under a moral sentimentalist framework. A liberal conception of justice allows us to criticize society's conventional rules as unjust when they are unfair, and our sympathy with

those unfairly victimized by society can lead us to demand social reform, or to practice civil disobedience if this reform is unduly delayed.

A liberal concern for fairness to individuals, however, is not a feature of Hume's own theory of justice, nor is it part of either the utilitarian conservative traditions that developed out of it. What all these positions have in common is the insensitivity to claims of the individual and to the propriety of the resentment an individual feels if sacrificed either to the public interest or to the dictates of traditional conventions. It is in this sense that contemporary liberals have generally been wary of Hume, afraid that the streams of thought growing from Hume's work could never provide a normative justification for a theory of justice focused on fairness to each particular individual.

5. The Overlooked Solutions

Yet political philosophers were mistaken about the fragilities of moral sentimentalism. It is a fine theory, and one that political philosophers now more than ever should seriously consider embracing. Hence, in this section, I am going to respectively address each of the problems elucidated in the previous section, showing how they can be solved within the boundaries of a moral sentimentalist account of human morality.

(i) *The Natural Fallacy Problem*

According to the *natural fallacy problem*, as discussed in the previous section, one cannot derive *ought* prescriptions from *is* descriptions. The reasons why this critique should not generally worry us when developing normative political theories has already been the object of debate in the second paper that forms this dissertation. Nonetheless it is still important to develop the argument for the immunity from this critique by moral sentimentalism more specifically. For it is one thing to claim that one can – and should – disregard the criticism endorsed by defenders of the natural fallacy and make use of empirical evidence in normative political theorizing, and yet another for one to stand for one particular strand of normative theory in the face of the same criticism.

In view of Hume's well-acknowledged inclinations towards naturalistic and psychological explanations of human morality, he has been widely accused of falling into an exclusively descriptive account of morality, devoid of any normative force. In making the case for moral sentimentalism in political philosophy it becomes imperative to deal with this accusation and to elucidate in which manner Hume's account of morality is not merely descriptive. Fortunately, Michael Frazer has successfully undertaken this task and I will here expand on his arguments.

The key to understanding the normative authority of a moral sentimentalist theory lies in the idea of reflective equilibrium. The term has been made popular by Rawls in his theory of justice and regards the exercise of reasoning back and forth from our considered judgments to our intuitions until we reach a point in which they are in sufficient consonance. At this point, we can say that we have achieved reflective equilibrium and, consequently, we can ascribe normative authority to the moral judgments that emerged from this process. In like manner,

> For Hume, what is important to consider about any given faculty is whether it can find a place in a mind that is reflectively stable, all things considered. (…) If our moral sentiments of the virtues of which they approve were to be rejected by our intellectual faculties as somehow false or unwarranted – or if they were to be rejected by our aesthetic faculties as somehow ugly or distasteful – then these too might be reasons for rejecting such sentiments in our search for reflective equilibrium. (Frazer, 2010, p. 59)

It is imperative to distinguish two aspects of Hume's reflective stability so as to be able to properly appreciate his normative account of moral sentiments. Firstly, there is the feeling of self-approbation that sentiments inspire towards themselves. This feeling is described as the dignity of virtue and constitutes a very limited form of reflective equilibrium. In this respect, Frazer stresses that

> In Korsgaard's terminology, the dignity of virtue is a matter of our moral sentiments successfully bearing the test of their own 'direct reflexivity.' As a result, the dignity of virtue alone is not sufficient to establish its normative authority. (…) Hume explicitly rejects what Korsgaard calls the theory of 'normativity as direct reflexivity.' If this account of normativity were the only one available to Hume, then this would be a very good case for the position that Hume's ethical project is a purely descriptive one. (Frazer, 2010, pp. 58-59)

Secondly, and most importantly, is the aforementioned reflective stability of the mind as whole. At the end of the *Treatise*, it is not on the self-approbation of individual mental faculties alone that Hume relies in order to explain the enforcing nature of morality. In his own words:

> The same system may help us to form a just notion of the *happiness*, as well as of the *dignity* of virtue, and may interest every principle of our nature in the embracing and cherishing that noble quality. Who indeed does feel an accession of alacrity in his pursuits of knowledge and ability of every kind, when he considers, that besides the advantages, which immediately result from these acquisitions, they also give him a new lustre in the eyes of mankind, and are universally attended with esteem and approbation? And who can think any advantages of fortune a sufficient compensations for the least breach of the *social* virtues, when he considers, that not only his character with regard to others, but also his peace and inward satisfaction entirely depend upon his strict observance of them; and that a mind will never be able to bear its own survey, that has been wanting in its part to mankind and society?
> (T 3.3.6.6)

In the above passage Hume clearly states that the reason for considering morality a strongly compelling guide for action relies in the broader necessity we have to be in peace with ourselves, which is only possible when our minds are able to bear their own surveys – as a whole. Hume is not concerned with the isolated approval of each of our faculties (moral, aesthetic, etc.), rather his focus is on a mind that is able to approve of itself when reflecting on itself *as a whole*. His reflective stability is a *holistic* reflective stability.

It is also important to note, at the end of the *Treatise*, the rationale that Hume offers for pursuing a holistic reflective equilibrium in the first place. He explicitly says that we strive to be able to bear our own survey not for its own sake, but for the 'peace and inward satisfaction' such stability provides (T 3.3.6.6). As has already been observed, Hume sees human beings as having a strong desire to avoid psychological contradictions from any source – both internal or external – in order to avoid the painful uneasiness created by such contradictions.

Although satisfaction with oneself cannot be said to be the whole of human happiness – which Hume repeatedly insists also requires a wide variety of other goods – it can certainly be thought to be a necessary part of it. This is why Hume devotes so much of his moral philosophy to emphasizing how a commitment to treat the

evaluations of our corrected moral sentiments as authoritative is wholly compatible not only with the quest for reflective equilibrium as such, but also with the self-interested pursuit of all the other variety of goods necessary for individual happiness.

Once we understand "what it would be like not to have a sense of justice – that it would be to lack part of our humanity too – we are led to accept our having this sentiment" (Rawls, 1971, pp. 428-29). Rawls's idea of this reflective self-acceptance is directly parallel to the mode of normative justification with which, in his lectures on moral philosophy, he portrays Hume's moral sentimentalism. Here, Rawls imagines a contemporary reader objecting to a sentimentalist ethics as nothing more than descriptive moral psychology. Yet to maintain such a position, Rawls counter arguments, is "seriously to misunderstand Hume."

Focusing on the conclusion of the *Treatise,* Rawls instead interprets Hume as maintaining "that his science of human nature (…) shows that our moral sense is *reflectively stable:* that is, that when we understand the basis of our moral sense – how it is connected with sympathy and the propensities of human nature, and the rest – we confirm it." This interpretation seems to be more related with the notion of reflective equilibrium than Rawls himself may have imagined. When we check over and again our moral intuitions against our considered judgments, we are unveiling a stable set of principles that are generated by the joint work of our rational *and* affective faculties. And it is from this stability (plus overlapping consensus) that the normative power of moral sentimentalist principles emanates.

(ii) *The Stability Problem*

In the *Treatise* Hume details the causes of the moral sentiments, providing a naturalistic psychological explanation for why agreeable and advantageous traits prove to be the ones that generate approval. He claims that the sentiments of moral approval and disapproval are caused by the operations of sympathy, which is not itself a feeling but rather a psychological mechanism that enables one person to receive by communication the sentiments of another – analogous to that which we would call empathy today.

Moral rationalists claim that moral properties are discovered through the use of our rational faculty alone. While they argue that all that is morally good is in accord with reason and all that is morally evil is unreasonable, Hume rejects both theses. Hume claims that our moral distinctions are not derived from reason but rather

from sentiment. On his distinct version of morality, the moral sentimentalist one, Hume interprets our moral beliefs as ideas copied from the impressions of approval or disapproval that represent a trait of character or an action as having whatever quality it is that one experiences in feeling the moral sentiment. Thus Hume's claim that moral good and evil are like heat, cold, and colors as understood in "modern philosophy," which are experienced directly by sensation, but about which we form beliefs.

We are able to determine that every trait, i.e., virtue, toward which we feel approval has at least one of the following four characteristics: it is either immediately agreeable to the person who has it or to others, or it is useful (advantageous over the longer term) to its possessor or to others. Vices prove to have parallel features: they are either immediately disagreeable or disadvantageous either to the person who has them or to others. These are not definitions of 'virtue' and 'vice' but empirical generalizations about the traits as first identified by their effects on the moral sentiments. As discussed in the previous section, this account of morality may lead to the instability of human moral judgments on two domains: diachronically within the same individual, and synchronically across individuals. Nonetheless, once again this criticism is the result of a misapprehension of Hume's moral theory, and such is the case for two different reasons.

Firstly, Hume does not understand our moral evaluations as purely a matter of sentiment. Quite the contrary, as he emphasizes, our capacity for sympathy includes a strong cognitive component. Hume's account of sympathy necessarily involves both an idea (rational) and an impression (sentimental) of the passion of its object of sympathy. Thus he is able to speak of rational and irrational types of sympathy and, with the use of this distinction, he is able to partially overcome the instability criticism. Sympathy is assumed to be irrational if it is either "founded on the supposition of the existence of objects which really do not exist" or "when in exerting any passion in action, we choose means insufficient for the designed end and deceive ourselves in our judgment of causes and effects" (T 2.3.3.6). In this sense, Frazer clarifies this first way in which Hume partially addresses the instability critique:

> It is clear that sympathy can go wrong in either or both of these ways. To use Philip Mercer's terminology, sympathy can be *misplaced* when we form mistaken ideas about another's state of mind, just as it can be *misguided* when we fail to take the appropriate means to the ends it suggests to us. In order to avoid these cognitive errors, sympathy must be accompanied by sound reason, both in inferring the (actual or hypothetical) passions of others and in deliberating about which actions will best achieve the ends that sharing these passions suggests to us. (Frazer, 2010, p. 42)

Secondly, the stability of a moral system can be undermined by the synchronic discrepancy of moral judgments across individuals. In this second sense, cross-individual moral stability can accomplished via commonsense and the practices of everyday life. As Hume clarifies, in everyday life we all believe some to be better judges than others and appeal, in matters of aesthetic and moral disagreement, to this 'higher and more refined taste, which enables us to judge of the characters of men, of compositions of genius, and of the productions of the nobler arts' ('Of the Delicacy of Taste and Passion,' in EMPL, p. 6). In order to achieve such a refined taste, we need to undergo a 'progress of the sentiments', moving beyond our immediate emotional reactions to phenomena. Unlike Rousseau, who argues that reason annihilates our natural feelings of *pitié*, Hume maintains that reason is a fundamental part of the process of refinement of our sympathy-derived moral sentiments (Frazer, 2010, p. 47, 48).

In this context, Hume stresses the crucial role played by a humanistic education in enabling the progress of our moral sentiments. Notwithstanding the innate character of our abilities of imagining, feeling, and reasoning about fine distinctions, we must not underestimate the impact that education has on improving these abilities. Through a liberal arts education, men

> (…) feel an increase of humanity … Thus *industry, knowledge,* and *humanity* are linked together by an indissoluble chain and are found, from experience as well as reason, to be peculiar to the more polished, and, what are commonly denominated, the more luxurious ages. ("Of Refinement in the Arts," in EMPL, p. 271)

Moreover, "Hume maintains that civilization and education can only help refine the sentiments, making them more sensitive to the feelings of our fellows – and

hence more suitable as grounds for a commitment to genuine justice and virtue" (Frazer, 2010, p. 48). There will still be no warranty that our tastes are always going to be in agreement, even after undergoing this process of refinement. In this sense, Hume accepts a certain level of diversity amid aesthetic judgments as a fact of the world (EMPL, p. 244). Nonetheless such disagreement is inacceptable for him in the moral sphere.

In order to overcome this difficulty and remain within the realm of sentimentalist morality, Hume avails himself of a device he terms 'the general point of view.' The idea of the general point of view is that people do not make moral judgments from their own individual points of view, but instead select "some common point of view, from which they might survey their object, and which might cause it to appear the same to all of them" (T 3.3.1.30). This device enables Hume to explain more generally how the moral evaluations made by one individual at different times and by distinct individuals at the same time tend to be fairly uniform.

In this sense, Hume agrees that the bias built into our sympathy is a proper ground for objecting to moral sentimentalism. While our sympathy is necessarily stronger for those to whom we have some connection, our moral evaluations, we feel, ought not to vary accordingly. Hence the general point of view aims at correcting this bias.

The adherence to this common point of view takes place when our moral sentiments have been refined and developed, such that we are able to reason about the best perspective to assume in order to engage in moral evaluation. Hume describes this process in the following passage:

> Our situation, with regard both to persons and things, is in continual fluctuation; and a man, that lies at a distance from us, may, in a little time, become a familiar acquaintance. Besides, every particular man has a peculiar position with regard to others; and 'tis impossible we cou'd ever converse together on any reasonable terms, were each of us to consider characters and persons, only as they appear from his peculiar point of view. In order, therefore, to prevent those continual *contradictions*, and arrive at a more *stable* judgment of things, we fix on some *steady* and *general* points of view; and always, in our thoughts, place ourselves in them, whatever may be our present situation. In like manner, external beauty is determined merely by pleasure; and 'tis evident, a beautiful countenance cannot give so much pleasure, when seen at the distance of twenty paces, as when it is brought nearer us. We say not, however, that it appears to us less beautiful: Because we know what effect it will have in such a position, and by that reflection we correct its momentary appearance. (T 3.3.1.15)

The ultimate relevance of this method for moral evaluation, Hume argues, is that life in society becomes impossible unless we share at the least some subset of standards of conduct and character to regulate our interactions. While we can coexist indefinitely amidst irresolvable aesthetic disagreement, we must have a common viewpoint for purposes of ethical judgment in order to live together at all. Although some degree of moral disagreement will still be inevitable, the existence of a shared moral viewpoint enables us to negotiate our social coexistence despite this disagreement, hopeful that our arguments will result in the achievement of real consensus.

Hume's arguments regarding the need for a general point of view are often presented exclusively in these social terms and the drive for moral consensus that they evoke is sometimes tied to Hume's political conservatism. Opposing this school of interpretation, Annette Baier (1988) writes that, for Hume, "the problem that corrected morality solves is deeper; it is as much intrapersonal as interpersonal," solving "contradictions in our individual sentiments over time" (p. 757). It is in Hume's analysis of the resolution of intrapersonal contradictions via the correction of our moral sentiments from a general point of view that the reformist potential of his ethics becomes most evident. It is precisely at this point that Hume clarifies why we can never fully exclude any of our fellows from the sphere of our moral concern (Frazer, 2010, p. 52).

Hume identifies a virtue he calls 'strength of mind' with the ability to act on calm, settled principles of action (T 2.3.3.10). The goal of moral development is to make the corrected moral sentiments the predominant inclinations in our souls, giving us the strength of mind to govern our moral evaluations, as well as our behavior, according to the principles of action that they provide (Frazer, 2010, p. 55).

To be sure, if we find that we lack such strength of mind, and our evaluations or behavior are determined by uncorrected moral sentiments or other violent passions, a mere will to change our psychological make-up is pointless (Frazer, 2010, p. 55). It is our sentiments, as Hume famously argued, which determine our actions, and 'the will never creates new sentiments' (T 3.2.5.5). We will have to begin by changing our actions, not our sentiments. Hume writes,

> But may not the sense of morality or duty produce an action, without any other motive? I answer, It may: But this is no objection to the present doctrine. When any virtuous motive or principle is common in human nature, a person, who feels his heart devoid of that principle, may hate himself upon that account, and may perform the action without the motive, from a certain sense of duty, in order to acquire by practice, that virtuous principle, or at least, to disguise to himself, as much as possible, his want of it. (…) But tho', on some occasions, a person may perform an action merely out of regard to its moral obligation, yet still this supposes in human nature some distinct principles, which are capable of producing the action, and whose moral beauty renders the action meritorious. (T 3.2.1.8)

Whenever we face strong opposition 'the efforts, which the mind makes to surmount the obstacle, excite the spirits and enliven the passion' (T 2.3.4.6), so the recalcitrance of our own uncorrected sentiments will only further spur us to rise to the challenge. Hume is optimistic that we will eventually find ourselves governed by refined, corrected moral sentiments (Frazer, 2010, pp. 55-56).

(iii) *The Problem of the Separateness of Persons* and *Hume's Conservatism Problem*[46]

For starters, as already discussed in the previous section, there are two separate but interrelated ways in which sympathy may be said to threaten the separateness of individuals: firstly, by eliminating the distinction between a sympathizer and the individual object of her feeling and, secondly, by eliminating the distinction between the multiple objects of a single person's sympathy. As stressed by Frazer (2010):

> Rawls acknowledges that sympathizing with multiple persons need not necessarily lead to an aggregate conception of general utility, but has trouble conceiving of any practical alternative. Sympathetic concern for multiple others, unless the interests of these persons are aggregated, would be thrown into confusion once the claims of these persons conflict. (p. 94)

Hume claims that the rules of justice are always in the interest of every individual; which, if true, would solve the problem. Yet Hume's claim that strict

[46] I am going to lump both these problems together under this subdivision due to the fact that their solutions rely on the one same improvement over Hume's theory; namely, Adam Smith's development of the impartial spectator.

justice is always in the interest of all individually cannot withstand reflective scrutiny. An adequate theory of justice must address what should be done when our interests conflict. The utilitarian solution to our conflicting interests is to aggregate the interests of the multiple objects of our sympathy – allowing benefits to some to outweigh harms to others. Yet this solution fails to address Rawls's criticism. Hence we are left with a real problem if we wish to stand for moral sentimentalism while at the same time embracing a liberal theory of justice capable of acknowledging the inviolability of persons.

In *The Theory of Moral Sentiments* Adam Smith provides the solution to this quandary. The first step of his solution involves the development of a different 'general point of view', which he calls the impartial spectator. In his own words:

> (…) in the same manner, we either approve or disapprove of our own conduct, according as we feel that, when we place ourselves in the situation of another man, and view it, as it were, with his eyes and from his station, we either can or cannot entirely enter into and sympathize with the sentiments and motives which influenced it.
> (TMS III. 1.2)

In the above description we can already detect an important aspect of Adam Smith's impartial spectator, namely, the fact that his spectator does not directly sympathize with the pleasure (or pain) felt by the individual under evaluation. There is an intermediate step in his process of sympathy. That is, we firstly judge the fitness of the individual's response (pleasure or pain) to his context before we sympathize with his feelings.

Additionally, Smith clarifies that he has no interest in what sort of moral sentiments might lead a "perfect being" to approve of a theory of justice, but rather "upon what principles so weak and imperfect a creature as man actually and in fact approves of it" (TMS II.5.10). As Frazer emphasizes, "The impartial spectator is an ideal type (…) in Max Weber's descriptive, sociological sense," not a perfect spectator as one could assume under some alternative perfectionist interpretation (2010, p. 95).

Whilst Hume uses the term 'sympathy' strictly as denoting a psychological mechanism, Smith embraces a much wider connotation for the same term. For Hume, sympathy is the process through which an idea of another's feeling is transformed into

an impression of it; for Smith, this account of sympathy is able to address only a small minority of our experiences of fellow feeling. Smith's concept of sympathy contains more cognitive elements than Hume's concept. Rather than simply copying others' feelings or thought processes as we imagine them, Smith's account of sympathy involves placing ourselves in the other's situation and working out what to feel, as though we were they. The degree of cognitive and imaginative effort required by a spectator will vary with what the actor being observed is experiencing. A spectator's sympathy may seem almost automatic when she is faced with strong, simple emotions such as sudden grief or joy, but will necessarily involve considerable imaginative effort with more complex and nuanced sentiments.

Smith provides us with a number of arguments in defense of his theory. Most importantly, he reminds us that while Hume's account of sympathy as emotional contagion may be a plausible description of *shared* pleasure it is not a plausible description of *sympathy with pain*. For Hume, "the minds of men are mirrors to one another, not only because they reflect each other's emotions, but also because those rays of passions, sentiments and opinions may be often reverberated, and may decay away by insensible degrees" (T 2.2.6.21).

This 'minds are mirrors' account of sympathy reveals a tendency of sympathetically shared feelings to reinforce each other. Hume gives an example of this tendency when he remarks that the pleasures that a rich man receives from his possessions will through sympathy cause pleasure in a sympathetic spectator; whose pleasure will in turn cause a new pleasure in the rich man. In Hume's words,

> (…) the pleasure, which a rich man receives from his possessions, being thrown upon the beholder, causes a pleasure and esteem; which sentiments again, being perceived and sympathized with, increase the pleasure of the possessor; and being once more reflected, become a new foundation for pleasure and esteem in the beholder. (T 2.2.6.21)

In contrast to this view, Smith argues that far from reinforcing our painful emotions, we find another's sympathy with our miseries comforting. As he explains, there is something about sympathy with another human being that is inherently pleasant to both parties involved, regardless of whether the feelings being shared are positive or negative. This observation, as Hume observed in a letter to Smith, is the 'hinge' of Smith's thought and one that Hume felt to be highly questionable. Hume

writes: “If all sympathy were agreeable, a hospital would be a more entertaining place than a ball” (CAS, Letter 36, p. 43; as in Frazer, 2010, p. 99).

Yet Smith does not succumb to Hume’s ironic critique; quite the contrary, he overcomes it by making use of his more sophisticated account of the mechanism of sympathy. Unlike Hume’s mechanism of emotional contagion, for Smith, sympathy works via a projective mechanism through which we only sympathize with the feelings of other individuals if we judge them as fit to their circumstances. Differently from the mechanism of emotional infection, the mechanism of projective empathy involves discretion on the part of the spectator as to whether sympathy is called for. As Darwall (1998) has observed,

> (…) we place ourselves in the other’s situation and work out what *to* feel, as though we were they. This puts us into a position to *second* the other’s feeling or dissent from it. As Smith puts it, we thereby express our sense of the “propriety” of the other’s feeling, whether, that is, we think it warranted or not. If we cannot ‘enter into’ an angry person’s sense of a situation that provokes her anger, we will feel her anger inappropriate (TMS.11). Or if a person laments his misfortunes, but ‘bringing [his] case home to ourselves’ does not affect us similarly, we will not share his grief but think it unwarranted (TMS.16). (p. 268)

The mechanism of sympathy under Smith’s account is constituted of a multiple stages. First, a spectator imaginatively engages with the situation of an actor, imagining what it would be like to be that actor in that situation. Second, the spectator feels some reaction herself in response to this imagined situation. The emotions she feels are akin to those she would feel if she were actually the actor in this situation, albeit without the vehemence that a real situation would inspire. The experience of this second stage of sympathy may or may not be pleasant and the imagined reactions of the spectator may or may not resemble the actual reactions of the actor. Indeed, the actor may not have any reaction to his situation at all, as when we sympathize with the deceased. Third, whenever possible, the spectator compares her reactions to the reactions of the actor, noting their degrees of similarity and difference. Fourth, there are the evaluations arising from this comparison – pleasurable approval to the extent that the actor’s and the spectator’s reactions to the situation in question resemble one another and painful disapproval to the extent that they do not.

Yet Smith, like Hume, realized that our sympathy varies along with the closeness of our relationship to the objects of our feelings and hence that our

judgments of propriety will be biased in favor of those closest to us. To avoid the social and psychological contradictions that result, an additional, fifth stage in the process of mature moral evaluation is required when we do not ourselves qualify as impartial spectators. In such cases, we correct our biased judgments through appeal to an imagined impartial spectator within, the functional equivalent of Hume's appeal to the general point of view.

It is now possible to see how Smith is able to escape from the first of the two pitfalls mentioned previously – namely, that of dissolving the distinction between a spectator and the object of her sympathy. Rawls would understand this weakening and modification to be a result of sympathy's imperfection. A perfect sympathizer, an ideal impartial spectator, would identify completely with the object of her sympathy. Yet it is essential to Smith's theory that sympathy can never be perfect in this way. If it were, the distance necessary for an appraisal of the actor's reactions to his situation would be impossible. The spectator cannot forget that she is a separate person from the actor, for she must contrast the actor's actual reactions to his situation with how she would react in his place. It is thus impossible for Smith to speak of 'perfect' sympathy'.

Smith is also able to simultaneously escape from the second pitfall, namely, that of eliminating the distinction between the multiple objects of a single person's sympathy, and from the threat of conservatism. Smith's solution to these problems relies on his interpretation of the concept of justice. He understands justice not as compliance with a set of rules, like Hume, but rather as a feature of persons – more specifically, as a virtue that allows an individual to avoid demerit. In this sense, Smith conceives justice as a negative virtue, something that is in place so that we can *avoid* hurting others. For him, "The sense of demerit is a compounded sentiment … made up of two distinct emotions; a direct antipathy to the sentiments of the agent and an indirect sympathy with the resentment of the sufferer" (TMS II.5.5). Well along in the TMS, Smith goes on to insist that the general rules of morality:

> (…) are ultimately founded upon experience of what, in particular instances, our moral faculties, our natural sense of merit and propriety, approve or disapprove of. (…) We do not originally approve or condemn particular actions because, upon examination, they appear to be agreeable or inconsistent with a certain general rule. The general rule, on the contrary, is formed by finding from experience that all actions of a certain kind, or circumstanced in a certain manner, are approved or disapproved of. (TMS III.4.8)

A polity may thus be called just to the degree that its positive law is a successful approximation of the natural law–that is, to the degree that its legal code accurately reflects the general rules of justice that may be inductively derived from particular impartial judgments. Unlike Hume's theory of justice, Smith's theory thus allows us to both properly account for the inviolability of the individual and to criticize an existing social order as unjust. Our proper and impartial sympathetic approval of the warranted resentment of those victimized by an unjust society leads us to demand social reform.

5.1 A Last Piece of the Case in favor of Moral Sentimentalism

Up to this point I have made the case for a moral sentimentalist turn in political philosophy based on two claims: (i) the empirical sciences have in the past few decades been providing accumulating evidence in favor of a sentimentalist understanding of human morality; and (ii) all the major arguments against moral sentimentalism are misplaced. Yet I would still like to present one more argument in defense of a sentimentalist approach for normative political theory.

This argument can be found in the literature regarding the possible implications, for meta-ethical theories, of the new findings from neuroscience and moral psychology. As explicated by Joyce (2008), "the general worry is that empirical discoveries about the genealogy of moral judgments may undermine their epistemic status and ultimately detract from their authoritative role in our practical deliberations." He goes on to add that "this is a possibility to be taken seriously and explored carefully" (p. 392). As I have discussed in the previous section, these empirical discoveries are not a real threat to the normative force of our moral commitments. Yet an interesting additional argument for moral sentimentalism has emerged from the debate spurred by this literature. In order to make sense of this argument, I will firstly have to expose part of this debate. At the end of the exposition, the argument will become clear.

The bulk of what is at stake is the plausibility of moral rationalism. And the main worry is: do the new results from neuroscience and moral psychology threaten the possibility of moral rationalism? Joyce (2008) argues that the answer is contingent on what researchers mean when they refer to 'moral rationalism'. He claims that

while some kinds of rationalism are indeed threatened by the recent findings, some other types remain immune. One first would have to clarify the concept, drawing the appropriate distinctions, so that one could then see which kinds of rationalism remain unchallenged by the empirical work. *Psychological* and *Conceptual Rationalism*, according to Joyce's reasoning, would indeed both be excluded from our moral scenery if recent empirical results are indeed solid and confirmed by future experiments.

Yet one important kind of rationalism, namely *Justificatory Rationalism*, would remain immune to all these novel empirical findings. Justificatory rationalists assert that moral transgressions amount to transgressions of rationality, having Peter Singer as one of their most prominent representatives. Their claim is distinct from the one made by the conceptual rationalist insofar as they do not assert any conceptual connection. Singer argues that natural selection has granted humans an innate tendency to look favorably upon actions that benefit one's family and a tendency to dislike actions that harm them. However, as Singer goes on to argue, we have also been granted by natural selection a rational faculty. Thus, we are able to reason our way out of parochialism. In his words, we transcend moral parochialism via the use of reason, reminding ourselves that:

> I am just one being among others, with interests and desires like others. I have a personal perspective on the world, from which my interests are at the front and center of the stage, the interests of my family and friends are close behind, and the interests of strangers are pushed to the back and sides. But reason enables me to see that others have similarly subjective perspectives, and that from 'the point of view of the universe' my perspective is no more privileged than theirs. (Singer, 1995, p. 229)

Based on Singer's argument, reason demands from us the recognition of an objective value in the welfare of others. According to Joyce, this constitutes the central intuition motivating many moral rationalists, such as Michael Smith, Christine Korsgaard, Thomas Nagel, Alan Gewirth, and even Kant. From a human psychological perspective, all that is necessary for justificatory rationalism to be reasonable is that persons fulfill the minimal requirements for being rational agents. This minimal constraint is not threatened by neuroscience, affirms Joyce (2008).

Nonetheless, Shaun Nichols contests Joyce's conclusion and, in doing so, offers an interesting argument in defense of a sentimentalist morality. Nichols (2008) claims that Singer's argument:

> (…) calls attention to the salient fact that from the perspective of the universe, there is no rational basis for privileging my own perspective. This leads the justificatory rationalist to the conclusion that we should reduce pain and suffer wherever we find it. But by what conveyance do we get to move from 'from the perspective of the universe there is no rational basis for privileging my own perspective' to 'rationality indicates that we should reduce pain and suffering, wherever it may be found'? Why, that is, should I give priority (or indeed credence) to the perspective of the universe when it comes to deciding the rational thing for me to do? (p. 401)

Nichols goes on to affirm that philosophers such as Peter Singer have been oblivious to a key factor in our moral reasoning: namely, the intuition that favoring my own perspective *seems* (or *feels*) unfair, unjust, and wrong. This claim is equivalent to stating that were we not to intuitively find the 'Justification Principle'[47] powerfully intuitive, we would be left with no possible purely rational case for its defense. The problem is that, if such a sentimentalist moral psychology is right, then it is probably illicit for the justificatory rationalist to rely on lay intuitions in favor of claims like the justification principle. Those lay intuitions are most likely a product of nonrational affective mechanisms, and it is quite possible that we would not find these claims intuitive if we lacked our affective responses.

A rationalist cannot rely on the intuitiveness of claims if their intuitiveness is rooted in our nonrational, emotional faculties. Thus justificatory rationalist arguments depend on intuitions that cannot carry the requisite weight if a sentimentalist account of those intuitions is correct – there cannot be a *rational* foundation for morality if this foundation is ultimately based on *affectively* charged intuitions.

6. The Possible Implications

In light of all the previous discussions, I will wrap up the case for embracing moral sentimentalism in political philosophy and I will suggest in a *rather incipient*

[47] The Justification Principle states that if my interests are not privileged form the perspective of the universe, then I should not privilege them to the exclusion of others' interests.

[48]manner some possible implications of such embracement. For instance, what would be the role of rhetoric in the political discourse? Would desert have a bigger role in theories of distributive justice?

In the preceding sections I have argued that together recent empirical findings from areas such as neuroscience and moral psychology make a strong case for moral sentimentalism. These empirical results, and its consequences, have already been acknowledged by many moral philosophers who have recently been paying due attention to sentimentalism in the development of their theories. In this context, my goal was to extend this acknowledgment to political philosophy, an area that is still to be illuminated by all these novel empirical findings.

In order to achieve this goal, I have first 'remade' the empirical case for moral sentimentalism. After exposing all the relevant empirical findings, I sided with Prinz (2006) in defending that emotions can directly cause moral evaluations and that, unlike conventional rules, moral rules are fundamentally grounded in emotions. On this sentimentalist view, believing that something is morally wrong is in essence having ''a sentiment of disapprobation'' towards it (Prinz, 2006, p. 33). In other words, condemning an act as immoral entails the experience of a negative emotional reaction, and the judgment itself is just an expression of this emotional reaction. In this sense, emotions[49] do not only partly constitute our moral judgments, but they are also necessary and sufficient for them. Through this we can more accurately portray our morality in a Smithian perspective.

The second step in my argument has been the contention that one of the main reasons why contemporary political philosophers have not yet properly acknowledged the relevance of moral sentimentalism is a misinterpretation of the theory. Due to this misinterpretation, political philosophers embraced rationalism as the proper way of conducing their agenda in political philosophy and, as a consequence of rationalism, they ended up alienated from all sorts of empirical data from a variety of distinct sciences. Hence, the current political philosophical literature is to this day mainly characterized by rationalist moral theories.

The third step of my argument has been to make the aforementioned misinterpretation explicit. Thus, I have exposed the central problems that political philosophers have mistakenly attributed to a moral sentimentalist approach to justice.

[48] The implications of a moral sentimentalist political philosophy are the subject of future work.
[49] It is important to keep in mind that our moral emotions entail a cognitive element.

These are the problems that have got in the away of a proper acknowledgment of the relevance of moral sentiments in political philosophy.

The fourth step has been the discussion of the solutions to these problems. I have therefore addressed all the aforementioned problems attached to moral sentimentalism, so as to show that they can be solved within the realm of a sentimentalist approach. The solutions involved the reinterpretation of some of Hume's ideas, and the refinement of Hume's philosophy as carried out by Adam Smith.

At last, I wrap up my case for moral sentimentalism in political philosophy, claiming that all recent empirical evidence pointing to a sentimentalist nature of our morality, plus the fact that the problems attributed to the theory are not apt, together constitute good enough reasons for a sentimentalist turn in our understanding of justice. Thus incorporating moral sentimentalism into political philosophical theorizing is one of the important and positive consequences of taking empirical evidence seriously.

There are several possible implications of a sentimentalist turn in political philosophy, and addressing them is the subject of my future work. Yet here I will very briefly mention some of these potential implications. For instance, one important consequence of embracing moral sentimentalism in political philosophy is the need to rethink the relevance of rhetorical discourse in the political debate. Frazer (2010) brings this implication out to our attention in his recent work, arguing that contemporary political philosophy has been failing to appreciate the crucial role played by the *form* of political discourse. One instance of the importance of the form of the discourse can be illustrated by role historically played by metaphors in the political life. Likewise, some researchers have already been arguing that most of the discourses that have led to real political change are replete with metaphors. As an example, Lakoff (1996) discusses Martin Luther King's "I Have a Dream" speech, emphasizing that he used a considerable number of metaphors so as to make it possible for white Americans to 'feel' the segregation and, as a result of this 'feeling', to endorse a legal change in the system.

Another possible implication of an emotional (and empirical) shift in political philosophy is the resurgence of principles of desert in theories of justice. If fairness is related to our sentiments of approbation (and disapprobation) towards actions and individuals, then the fact that these sentiments are present in certain judgments of

desert should be ignored. Unquestionably much more philosophical and empirical work is needed to ascertain the extent to which principles of desert are relevant, and the kinds of desert that should be taken into consideration.[50] These implications, and others not mentioned here, have to be further discussed and developed by philosophers and empirical scientists. In this paper, I have solely intended to make the case for moral sentimentalism in political philosophy, and I hope to have succeeded.

[50] In the fourth paper I hope to have contributed to this debate.

Luck, Desert, and Fairness: An Empirical Investigation

Introduction

While John Rawls famously makes frequent reference to common sense morality and the normative relevance of our "everyday judgments," he is not alone in this regard among political philosophers. More recently, David Miller has similarly claimed that "a theory of justice brings out the deep structure of everyday beliefs that, on the surface, are to some degree ambiguous, confused, and contradictory" (2003, p. 51). On this view, developing a theory of justice involves the reflective interplay between moral principles and considered judgments – with ordinary moral intuitions serving as the starting point of one's philosophical investigations. However, as Miller points out, appeals to common sense morality are at least partly *empirical claims*. As such, philosophers cannot simply assume from the armchair that their *own intuitions* are representative of the beliefs and attitudes of the non-philosophical masses. This cautionary principle is one of the motivating forces behind the nascent but growing field of experimental philosophy.

From the outset, one of the central goals of experimental philosophers has been to shed empirical light on the contours of people's intuitions, beliefs, and attitudes about a wide variety of philosophical issues. Just in the past few years, philosophers have carried out experimental work in areas as diverse as epistemology, action theory, free will, ethics, philosophy of language, philosophy of law, philosophy of mind, and philosophy of science.[51] In each of these fields, a voluminous amount of work has been done in a relatively short amount of time. While the focus, methods, and goals of individual experimental philosophers vary widely, their interdisciplinary research is animated by a shared commitment (a) to using controlled and systematic experiments to explore people's intuitions, attitudes, and behaviors, and (b) to examining how the results of these experiments bear on traditional philosophical debates.

Perhaps unsurprisingly, not all philosophers agree when it comes to the relevance of the empirical findings that have been collected thus far (see, e.g.,

[51] For an overview of the various goals and methods adopted by experimental philosophers see Knobe (2007); Knobe & Nichols (2007); and Nadelhoffer & Nahmias (2007).

Ichikawa, 2011; Liao, 2008; Sosa, 2007; Williamson, 2011). However, when it comes to philosophical projects which seem to hinge at least in part on common sense morality, it's much more difficult to see how to justify the view that data on people's actual moral beliefs are philosophically irrelevant. Presumably, this explains why so much experimental work has been done on folk intuitions about free will and moral responsibility since these are areas where appeals to what laypersons think are commonplace.[52] Yet, at the same time, very little work has been done by experimental philosophers on political philosophy – an inattention that is especially curious given the role that claims about common sense morality often play in political philosophy.

One area where data on folk intuitions seem particularly germane is the longstanding debate about the complex relationship between luck, fairness, and desert. As Rawls makes clear in *A Theory of Justice*, according to one popular and influential strand of political thought, brute luck – e.g., being lucky in the so-called "lottery of life" – ought to have no place in a theory of distributive justice. On this view, it is a dictate of common sense morality that people can't properly be said to deserve to be rewarded (or punished) simply because they happen to be genetically advantaged (or disadvantaged). Similarly, fortuitous social circumstances such as wealth inheritance or family station and stability seem completely arbitrary from the standpoint of desert, fairness and justice. As Rawls stresses, "Intuitively, the most obvious injustice of the system of natural liberty is that it permits distributive shares to be improperly influenced by these factors [i.e., brute luck] so arbitrary from a moral point of view" (Rawls, 1971, p. 71).

According to egalitarian theories of distributive justice, while we might rightly admire someone for her natural abilities or her social station in life, these kinds of natural and social luck are not proper bases for judgments concerning desert or the proper distribution of resources. As Anthony Kronman claims, "It is unfair that people's fate should be determined, to a considerable degree, by a natural lottery" (1981, p. 76). But what evidence is there that common sense morality supports these types of egalitarian claims about the relationship between brute luck, desert, fairness, and distributive justice? Traditionally, these claims have been unmoored from any hard data concerning people's moral and political beliefs and attitudes. Instead, political philosophers have often simply assumed that their own beliefs reflect the

[52] For an overview of the work that has been done by experimental philosophers on intuitions about free will and related concepts, see Sommers (2011).

beliefs of the pre-theoretical masses.[53] So, while philosophers like Rawls and Kronman may turn out to be right about how laypersons think about luck, desert, and fairness, they may also turn out to be wrong. Figuring out which is the case is not something we can do from the armchair.

In this sense, I side with David Miller when he affirms that, "empirical evidence should play a significant role in justifying a normative theory of justice, or to put it another way, that such a theory is to be tested, in part, by its correspondence with our evidence concerning everyday beliefs about justice" (Miller, 2003, p. 51). In the present paper, I present the results of my own attempts to fill in some of the missing empirical details. But first, in the second section following this introduction, I set the stage with a discussion of the recent work in experimental political philosophy by Christopher Freiman and Shaun Nichols on folk intuitions about luck, desert, and fairness. Then, in the third section, I present my own attempts to build upon their work by correcting for some of the methodological shortcomings and limitations of their research. Finally, I conclude with a discussion of the relevance of my findings to political philosophy and I consider some possible future avenues of research. As we will see, there is a lot of interdisciplinary work that remains to be done.

2. Setting the Stage

Freiman and Nichols take as their starting point the assumption that philosophers "cannot simply assume that our intuitions are representative of the intuitions of laypersons, or even other philosophers" (2011, p. 124). In this respect, they share David Miller's aforementioned view that it is not enough to speculate about the distribution of intuitions among the folk. To the extent that political philosophers like Rawls are going to make claims about how people ordinarily think about justice and related concepts, they must take care to ensure these claims enjoy empirical support. For instance, political philosophers of an egalitarian bent have consciously avoided placing much stock in desert while at the same time suggesting that theories of justice are designed at least in part to describe "our sense of justice" (Rawls, 1971, p. 41). According to Scheffler (1992), this puts much of contemporary

[53] It is important to clarify that I do by no means intend to deny that Rawls appeals to empirical data in his philosophical work. My claim here is that Rawls, and other contemporary political philosophers, rarely (if ever) appeal specifically to data about commonsense morality.

political philosophy in conflict with ordinary morality – which purportedly gives pride of place to desert-related issues. Of course, this, too, is a descriptive claim about folk intuitions and practices that requires empirical support – but more on that later.

According to Freiman & Nichols, the main reason why desert has fallen out of favor with a number of contemporary political philosophers is that many of these philosophers adopt the so-called "brute luck constraint" – that is, the view that "if differential benefits are distributed on the basis of desert, brute luck cannot differentially affect the desert base (i.e., that which grounds the desert claim)" (2011, p. 124). The notion of brute luck operating here is to be distinguished from what Ronald Dworkin calls "option luck." As he says, "option luck is a matter of how deliberate and calculated gambles turn out – whether someone gains or loses through accepting an isolated risk he or she should have anticipated and might have declined" (Dworkin, 2002, p. 73). In short, option luck, whether good or bad, is something agents open themselves up to as the result of their deliberate and voluntary behavior (even if it's not something agents control in some deeper sense). Brute luck, on the other hand, is not a matter of voluntary or deliberate behavior. Instead, brute luck is something over which the agent has no control and for which the agent (supposedly) bears no responsibility.

So, while it may make some sense to say that an agent deserves more or less based on option luck, it makes little sense to say that an agent could deserve more or less based merely on brute luck – that is, brute luck ought not affect or influence one's desert base. As Wojciech Sadurski puts it, "We cannot, morally speaking, claim any credit for benefiting from circumstances which we have not brought about (…) through our conscious or deliberate actions. *That much is often accepted as a moral truth so obvious as not requiring any further defense*" (2007, p. 1, *emphasis added*). This is the reasoning behind the aforementioned brute luck constraint identified by Freiman and Nichols.

In this context consider, for instance, the following representative remark from Rawls: "We do not deserve our place in the distribution of native endowments, any more than we deserve our initial starting place in society" (1971, p. 89). This becomes all the more problematic once we acknowledge that we have no feasible method for ascertaining the extent to which a person's conscious efforts are attributable to his or her virtuous character rather than being attributable to valuable yet nevertheless undeserved natural abilities. In this sense, liberal egalitarian political

philosophers claim that the joint combination of natural and social luck ultimately undermines all claims of genuine desert. Accordingly, Rawls concludes, "the idea of rewarding desert is impracticable" (1971, p. 312).

It is partly for this reason that Rawls thinks that the traditional liberal conception of distributive justice ought to be rejected. As he says:

> While the liberal conception seems clearly preferable to the system of natural liberty, intuitively it still appears defective. For one thing, even if it works to perfection in eliminating the influence of social contingencies, it still permits the distribution of wealth and income to be determined by the natural distribution of abilities and talents. Within the limits allowed by the background arrangements, distributive shares are decided by the outcome of the natural lottery; and this outcome is arbitrary from a moral perspective. There is no more reason to permit the distribution of income and wealth to be settled by the distribution of natural assets than by historical and social fortune. (1971, pp. 63-64)

Along the same lines, Rawls goes on to claim that:

> The extent to which natural capacities develop and reach fruition is affected by all kinds of social conditions and class attitudes. Even the willingness to make an effort, to try, and so to be deserving in the ordinary sense is itself dependent upon happy family and social circumstances. It is impossible in practice to secure equal chances of achievement and culture for those similarly endowed, and therefore we may want to adopt a principle which recognizes this fact and also mitigates the arbitrary effects of the natural lottery itself. (1971, p. 64)

On this view, because an individual's social luck affects his prospects of development of his natural abilities, it becomes practically impossible to disentangle the extent to which his good or bad fortune is due to his effort (or lack thereof). Owing to this and related difficulties, Rawls concludes that the deeply intertwined notions of luck and desert have no proper role to play in an adequate theory of distributive justice. On this view, because both natural and social luck serve as practical impediments to ascertaining what people genuinely deserve, the ideal distribution of resources has to be based on something else.

In short, because Rawls adopts the brute luck constraint when it comes to desert and because he thinks "luck swallows everything," to borrow a phrase from Strawson (1998), Rawls doesn't think desert can play any role in an adequate theory

of distributive justice. Of these three claims, it appears that Rawls only attributes the brute luck constraint to common sense morality. Once again, in Rawls' words:

> Perhaps some will think that the person with greater natural endowments deserves those assets and the superior character that made their development possible. Because he is more worthy in this sense, he deserves the greater advantages that he could achieve with them. This view, however, is surely incorrect. *It seems to be one of the fixed points of our considered judgments that no one deserves his place in the distribution of native endowments, any more than one deserves one's initial starting place in society.* The assertion that a man deserves the superior character that enables him to make the effort to cultivate his abilities is equally problematic; for his character depends in large part upon fortunate family and social circumstances for which he can claim no credit. The notion of desert seems not to apply to these cases. (1971, pp. 103-104, *emphasis added*)

So, while the people on the street may not appreciate the ubiquitous role that luck plays in our lives, Rawls nevertheless thinks that most people realize that luck isn't a proper basis for claims of desert. However, not all philosophers agree that this is the ordinary view.

For instance, David Hume famously claims that commonsense judgments of desert are *insensitive* to whether an agent's successes or failures are a matter of option luck or brute luck. As he says:

> No distinction is more usual in all systems of ethics, than that betwixt natural abilities and moral virtues; where the former are plac'd on the same footing with bodily endowments, and are suppos'd to have no merit or worth annex'd to them. Whoever considers the matter accurately, will find, that a dispute upon this head wou'd be merely a dispute of words, and that tho' these qualities are not altogether of the same kind, yet they agree in the most material circumstances. They are both of them equally mental qualities: And both of them equally produce pleasure; and have of course an equal tendency to procure the love and esteem of mankind … Tho' we refuse to natural abilities the title of virtues, we must allow, that they procure the love and esteem of mankind; that they give a new luster to the other virtues; and that a man possess'd of them is much more entitled to our good-will and services, than one entirely devoid of them.
> (T 3.3.4.1)

David Miller sides with Hume when it comes to the brute luck constraint and common sense morality. On his view, folk attributions of desert are often blind to worries about luck and the natural lottery of life. As Miller says:

> If we consider the attitudes of admiration, approval, etc., it is plain that we do not adopt them only towards qualities believed to be voluntarily acquired. When we admire the superlative skill of a musician, we do not ask about the conduct which led to its acquisition before granting our admiration. The attitude is held directly towards the quality as it now exists, and the question, 'voluntarily acquired or not?' is simply not considered. If the close relation between appraising attitudes and desert is admitted, it seems inconceivable that such judgments as 'Green (the musician) deserves recognition' should not be made on the same basis: on the basis of the skill alone, without reference to the manner of its acquisition. And this is indeed our practice. (1976, p. 96)

Miller cites findings that seem to suggest that laypersons endorse differential desert claims based on agents' differential contributions.[54] And given that he thinks one's theory of desert ought to be empirically grounded in what we know about commonsense morality, he ends up partially endorsing a contribution theory of desert whereby agents deserve the fruits of their labors regardless of whether these fruits were ultimately the result of nothing more than brute luck. On this view, it is effort and the end product that serve as the desert base – that luck may have played a role is simply beside the point. Of course, this, too, is an empirical claim – a claim that Freiman and Nichols decided to put to the experimental test.

In order to explore whether folk intuitions about desert seem to be shaped by something like the brute luck constraint, Freiman and Nichols ran a series of simple experiments. In the first study, they presented participants with the following statement: "Suppose that some people make more money than others solely because they have genetic advantages" (2011, p. 127). Participants were then asked whether they thought these genetically advantaged people deserved extra money and whether they thought it was fair that these people received more money. The results conformed to the brute luck constraint. As Freiman and Nichols report, "On average, people maintained that the people who made more solely because of genetic advantages did not deserve the extra money, nor was it fair that they get the extra money" (2011, p. 127). However, while these preliminary findings support they claim

[54] See Miller, 2003, Chapter Four.

that the brute luck constraint is part of common sense morality, Freiman and Nichols thought that perhaps people would have different intuitions if they were presented with concrete rather than abstract cases.

In an effort to see whether "different kinds of judgments would manifest if people were presented with questions about concrete individuals," Freiman and Nichols ran two additional studies (2011, p. 127). After all, as they point out, there is a growing body of research in both social psychology and experimental philosophy which suggests that people's moral intuitions are influenced by how abstractly cases are described. Because desert and fairness are inherently moral concepts, Freiman and Nichols expected that they might find similarly asymmetrical results if they presented participants with concrete cases involving brute luck. So, they designed the following two concrete cases:

> Case 1: The Singers
>
> Suppose that Amy and Beth both want to be professional jazz singers. They both practice singing equally hard. Although jazz singing is the greatest natural talent of both Amy and Beth, Beth's vocal range and articulation is naturally better than Amy's because of differences in their genetics. Solely as a result of this genetic advantage, Beth's singing is much more impressive. As a result, Beth attracts bigger audiences and hence gets more money than Amy.
>
> Case 2: The Jugglers
>
> Suppose that Al and Bill both want to be professional jugglers. They both practice Bill's hand-eye coordination is naturally better than Al's because of differences in their genetics. Solely as a result of this genetic advantage, Bill can perform more difficult and impressive tricks than Al. As a result, Bill gets bigger audiences and hence more money than Al.

Participants in this follow up study were then presented with either the aforementioned abstract case or with one of these two concrete cases. They were then once again asked whether the genetically gifted individual(s) deserved the extra money and whether it was fair that they received the extra money. As predicted, Freiman & Nichols found statistically significant differences between how participants responded in the abstract and concrete cases:

> The statistical details are as follows. In the abstract condition, the mean response to the desert question was 2.78; the mean response to the fairness question was 2.72 (4 is the midpoint between "strongly disagree" and "strongly agree"). In the concrete singers case, the mean response to the desert question was 4.57, and the mean response to fairness was 4.71; for concrete jugglers, the mean responses for desert and fairness were 4.86 and 5.5. The differences between abstract and concrete were significant in all cases. People agreed more strongly with the claim that the singer deserved the extra money ($t(30)=2.51$, $p<.05$) and that the juggler deserved the extra money ($t(30)=3.09$, $p<.01$). Similarly, people agreed more strongly with the claim that it was fair for the singer to get the extra money ($t(30)=2.89$, $p<.01$) and also that it was fair for the jugglers to get more money ($t(30)=4.74$, $p<.001$).
> (2011, p. 129)

As Freiman & Nichols point out, these results suggest that people's moral intuitions about desert and fairness are influenced by the abstractness (or concreteness) of the case. As they say, "When faced with an abstract question, people's judgments conform to the brute luck constraint; when given concrete scenarios, people's judgments flout the brute luck constraint" (2011, p. 129). Having found an asymmetry in people's responses to the abstract and concrete scenarios, Freiman and Nichols go on to consider which intuitions "should guide our theorizing about justice" (2011, p. 129).

However, because I think that their studies have some methodological and conceptual shortcomings and limitations, I am going to postpone examining what they have to say on this front. So, while I applaud Freiman and Nichols's efforts to shed some empirical light on the debate in political philosophy about the relevance of brute luck to considerations of desert, fairness, and justice, I nevertheless think more experimental work needs to be done before we are able to consider the relative merits of concrete and abstract moral intuitions. In the following section, I am going to discuss my own attempts to fill in some of the missing details.

3. New Studies

Before I discuss my own attempts to explore folk intuitions about luck, desert, and fairness, I first want to highlight what I take to be the main shortcomings and limitations of the otherwise important work done by Freiman and Nichols. The first

problem with their studies concerns an important difference between their abstract and concrete conditions. In their concrete conditions, it is clear that (i) both individuals are equally hard working, and (ii) the agent is genetically advantaged *in the right way*, i.e., the genetic advantage is directly related to the performance of his or her work. After all, in both the singer and juggler cases, we are told that the two individuals are equally hard working and the genetic advantage that one of the agents have is completely specified. The only difference between these hard working agents is that the agent who earns more is genetically advantaged *in the right way*. This makes it clear that the agents who make more do so solely because they are genetically advantaged for the performance of that kind of work.

In the abstract condition, on the other hand, the case is described in such a way that leaves it open (i) whether the agent is hard working in addition to being genetically advantaged, and (ii) what is the kind of genetic advantage enjoyed by the agent. So, while the case does specify that the agents make more "solely because they have genetic advantages," this leaves open the possibility that these genetically advantaged individuals are making more money either (i) despite not working hard at all, or (ii) as a consequence of a genetic advantage of a sort unrelated to the type of work being performed by the agent. Because the concrete condition precludes this reading by specifying that the genetically advantaged are also hard working and genetically advantaged *in the right way*, it is worrisome that this difference may be driving the findings by Freiman & Nichols. Hence, I intend to address this worry in my studies.

There are two additional limitations in the studies ran by Freiman & Nichols. First, their studies only looked at natural luck even though social luck is another issue that looms large in the debates in political philosophy about desert, fairness, and justice. After all, brute luck isn't just limited to genetic advantages. One can be lucky in terms of social station and status as well – which is another issue I want to explore. Second, Freiman & Nichols only focused on the genetically advantaged. Yet I think it is equally important to probe people's intuitions about the genetically and socially disadvantaged as well. So, this is something I will address as well.

Thus the overarching goal of my present study is to advance the debate concerning the role played by the brute luck constraint in common sense morality. In my efforts to accomplish this goal, I created a series of new vignettes that were

specifically designed to address the aforementioned shortcomings and limitations of the otherwise important work done by Freiman & Nichols (see below).

To begin, I uploaded the study to Qualtrics.com – which is where the data was collected and stored. Participants were 404 people recruited via Amazon.com's Mechanical Turk online survey service and paid $1 each for completing the survey. Participants had to be at least 18 years of age and living in the United States. Fifty-one percent of participants (n = 204) reported being male, 49% (n = 200) reported being female. Eighty percent of participants (n = 322) reported being White/Caucasion, 11% (n = 43) reported being Black/African American, 5% (n = 19) reported being Asian/Pacific Islander, 3% (n = 12) reported being Hispanic/Latino, and 1% (n = 5) reported being Native American/American Indian. Participants' age ranged from 19 to 73, with a mean age of 38.84 (SD = 12.89). Age 29, 36, and 49 marked the 1st, 2nd, and 3rd quartiles, respectively. Finally, eighteen percent (n = 132) of participants reported an income of $0-20,000, 16% (n = 116) reported an income of $21,000-$40,000, 9% (n = 82) reported an income of $41,000-60,000, 5% (n = 37) reported an income of $61,000-80,000, 1% (n = 12) reported an income between $81,000-100,000, and 2% (n = 22) reported an income of more than $100,000.

Each participant was randomly assigned to one of four conditions: (a) Abstract positive, (b) Abstract Negative, (c) Concrete Positive, or (d) Concrete Negative. Each condition included four cases and each case was followed by a statement about desert and a statement about fairness. The two abstract conditions were as follows:

Abstract Positive:

Case 1

Suppose that some hardworking singers make more money than other equally hardworking singers solely because they have genetic advantages that make them artistically more talented.

Case 2

Suppose that some hardworking scientists make more money than other equally hardworking scientists solely because they have genetic advantages that make them intellectually more talented.

Case 3

Suppose that some hardworking athletes make more money than other equally hardworking athletes solely because they have genetic advantages that make them physically more talented.

Case 4

Suppose that some hardworking people make more money than other equally hardworking people solely because they had social advantages like a loving family and better education that made them more likely to be successful.

Abstract Negative:

Case 1

Suppose that some hardworking singers make less money than other equally hardworking singers solely because they have genetic disadvantages that make them artistically less talented.

Case 2

Suppose that some hardworking scientists make less money than other equally hardworking scientists solely because they have genetic disadvantages that make them intellectually less talented.

Case 3

Suppose that some hardworking athletes make less money than other equally hardworking athletes solely because they have genetic disadvantages that make them physically less talented.

Case 4

Suppose that some hardworking people make less money than other equally hardworking people solely because they had social disadvantages like an abusive family and worse education that made them less likely to be successful.

After each one of the abstract cases, participants were presented with the following two statements and asked to note their level of agreement on a 7-point Likert scale – ranging from 1 (strongly disagree) to 7 (strongly agree), with 4 as the midpoint (neither agree nor disagree).

1. These genetically advantaged [disadvantaged] singers [or scientists or athletes or people] deserve to make more [less] money.

2. It is fair [unfair] that these singers/scientists/athletes/people get more [less] money only because they have genetic advantages [disadvantages] that make them more [less] talented.

In the concrete conditions, participants received one of the following sets of four cases:

Concrete Positive:

Case 1

Suppose that Amy and Beth both want to be professional jazz singers. They both practice singing equally hard. Although jazz singing is the greatest natural talent of both Amy and Beth, Beth's vocal range and articulation is naturally better than Amy's because of differences in their genetics. Solely as a result of this genetic advantage, Beth's singing is much more impressive. As a result, Beth attracts bigger audiences and hence gets more money than Amy.

Case 2

Suppose that Amy and Beth both want to be software programmers. They both study equally hard. Although math is the greatest natural talent of both Amy and Beth, Beth is always able to come up with more efficient solutions for software programs than Amy because of differences in their genetics. Solely as a result of this genetic advantage, Beth's programming is much more impressive. As a result, Beth gets a better job and hence makes more money than Amy.

Case 3

Suppose that Amy and Beth both want to be professional basketball players. They both train equally hard. Although basketball is the greatest natural talent of both Amy and Beth, Beth is always able to naturally come up with better plays than Amy because of differences in their genetics. Solely as a result of this genetic advantage, Beth's playing is much more impressive. As a result, Beth gets a better position in a better team and hence makes more money than Amy.

Case 4

Suppose that Amy and Beth both want to be architects. They both study equally hard. Beth's parents are richer and able to pay for her to attend both a better school and a better university than the ones Amy's parents are able to afford their daughter. Solely as a result of this social advantage, Beth's curriculum is much more impressive. As a result, Beth gets a better job and hence makes more money than Amy.

Concrete Negative:

Case 1

Suppose that Amy and Beth both want to be professional jazz singers. They both practice singing equally hard. Although jazz singing is the greatest natural talent of both Amy and Beth, Beth's vocal range and articulation is naturally worse than Amy's because of differences in their genetics. Solely as a result of this genetic disadvantage, Beth's singing is much less impressive. As a result, Beth attracts smaller audiences and hence gets less money than Amy.

Case 2

Suppose that Amy and Beth both want to be software programmers. They both study equally hard. Although math is the greatest natural talent of both Amy and Beth, Beth is always able to come up with less efficient solutions for software programs than Amy because of differences in their genetics. Solely as a result of this genetic disadvantage, Beth's programming is much less impressive. As a result, Beth gets a worse job and hence makes less money than Amy.

Case 3

Suppose that Amy and Beth both want to be professional basketball players. They both train equally hard. Although basketball is the greatest natural talent of both Amy and Beth, Beth is always able to naturally come up with worse plays than Amy because of differences in their genetics. Solely as a result of this genetic disadvantage, Beth's playing is much less impressive. As a result, Beth gets a worse position in a worse team and hence makes less money than Amy.

Case 4

Suppose that Amy and Beth both want to be architects. They both study equally hard. Beth's parents are poorer and able to pay for her to attend both a worse school and a worse university than the ones Amy's parents are able to afford their daughter. Solely as a result of this social disadvantage, Beth's curriculum is much less impressive. As a result, Beth gets a worse job and hence makes less money than Amy.

After each one of the concrete cases, participants were presented with the following two statements and asked to note their level of agreement on a 7-point Likert scale – ranging from 1 (strongly disagree) to 7 (strongly agree), with 4 as the midpoint (neither agree nor disagree).

1. Beth deserves to make more [less] money.
2. It is fair that Beth gets more [less] money only because she has genetic advantages [disadvantages].

Before discussing the results, I want to briefly remind the reader of what I am trying to accomplish with this study. Firstly, I wanted to make it clear to participants that the advantaged (or disadvantaged) work equally as hard as their counterparts who make less (or more) money. Secondly, I also wanted to make it clear that the genetically advantaged agent was so *in the right way*. Thirdly, I wanted to explore people's intuitions about both genetic and social advantages (and disadvantages). Finally, I wanted to explore people's intuitions about negative luck and not just positive luck. As we will now see, each of these three factors yielded interesting results.

For starters, I found no differences within any of the four conditions between the three types of genetic luck – e.g., people treated being genetically advantaged at singing no differently than being genetically advantaged at science or sport. So, for the purposes of data analysis, I simply collapsed the three types of genetic luck into one variable. This left me with the following three factors: (a) genetic vs. social, (b) concrete vs. abstract, and (c) positive vs. negative. For present purposes, I decided to run a mixed factor ANOVA. The overall findings can be seen in Figure 1.

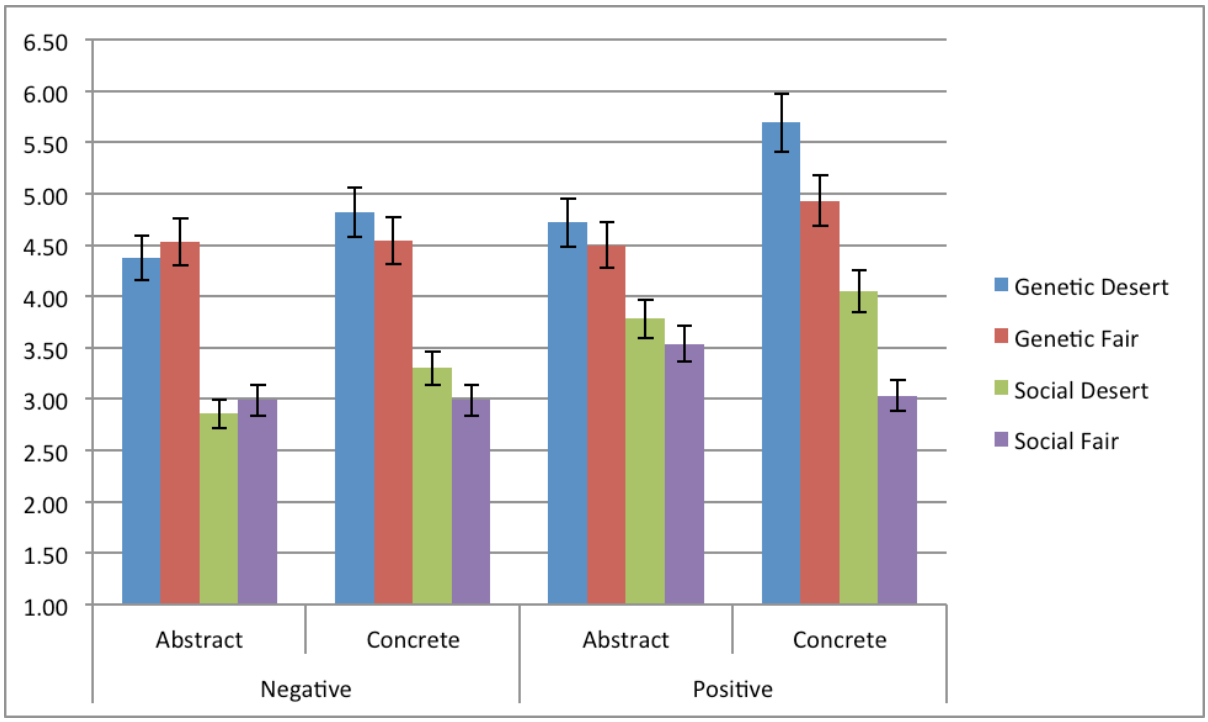

Figure 1: Overall Findings

The first thing worth pointing out is that I did not find significant within subject differences overall between participants' judgments about desert (M = 4.199, SD = 0.66) and their judgments about fairness (M = 3.881, SD = .069). However, there was one exception to this general trend – namely, judgments about desert and fairness did significantly come apart in the Positive Concrete condition, interestingly both for social and genetic luck (see Figure 1). This exception was an unexpected finding that certainly requires more research, and for which I will present some rather tentative hypothesis in the general discussion section.

The second noteworthy finding is that, considering the positive cases, people's judgments about desert and fairness were barely significantly sensitive ($p < .046$) to whether the cases were presented abstractly (M = 3.911, SD = .090) rather than concretely (M = 4.169, SD = .092). Most importantly, I found no shift in their judgments about desert from the concrete positive to the abstract positive cases. That is, participants judged the genetically advantaged agents as deserving of the additional money both in the abstract and the concrete positive cases; and they judged the socially advantaged agents as *not* deserving of the additional money both in the abstract and the concrete positive cases. This important finding reveals that the changes made to the experiment designed by Freiman & Nichols indeed affected the answers provided by participants – as I predicted they would.

The third remarkable finding is that there was a significant within subject difference ($p < .001$) between participants' judgments in response to genetic luck ($M = 4.763$, $SD = .072$) and their responses to social luck ($M = 3.317$, $SD = .080$). This suggests that common sense morality does not treat all kinds of brute luck equally. Instead, people tend to find social advantages and disadvantages to be much more problematic than genetic advantages and disadvantages. As I will discuss in the next section, this difference is to be expected if one endorses a moral sentimentalist perspective on desert.

At last, I found that people's judgments about desert and fairness significantly ($p < .001$) differ when the cases are positive ($M = 4.280$, $SD = .090$) rather than negative ($M = 3.8$, $SD = .090$). So, while people tend to think that advantages (both genetic and social) are deserved and fair, they find disadvantages to be more problematic. Interestingly, this different barely reaches significance in the assessment of fairness (see Figure 2).

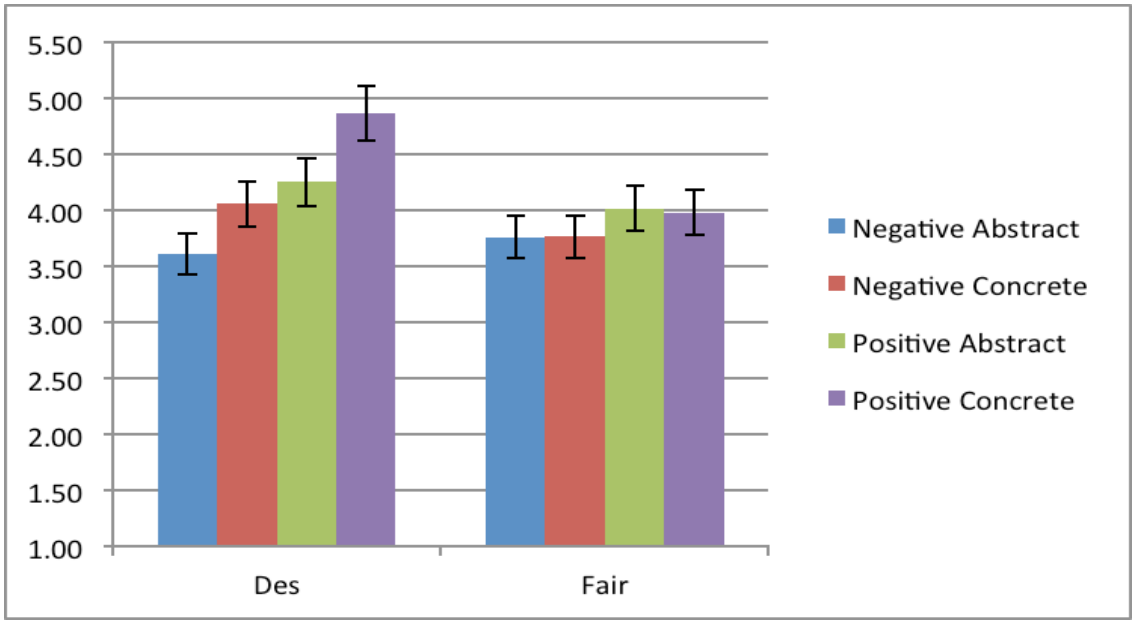

Figure 2: Desert versus Fairness

4. General Discussion and Future Directions

The debate about luck, desert, and fairness in contemporary political philosophy has just recently been rekindled by a handful of philosophers who claim that desert should play a bigger role in theories of distributive justice. I hope that my

experiment and its results will help to expand on this debate by shedding some light into the nuances of the folk intuitions about the concept of desert and its relation to different sources of luck.

As already discussed in the preceding sections, there are three distinct claims that one can make about desert and luck in theories of justice. The first claim is the so-called brute luck constraint, and considers neither natural nor social luck as proper grounds for desert. The second claim entails the pervasive presence of luck in our lives, and assumes two different forms: (i) luck swallows everything, following Strawson; or (ii) following Rawls, it is at the very least rather hard (not to say impossible) to disentangle that which was generated by brute luck (not voluntary) from that which was the result of a virtuous character or of individual effort (as much as possible considered to be voluntary). The third and final claim necessarily results from the concomitant assumption of the two former ones, stating the denial of any role for desert in theories of justice.

Contemporary political philosophers of the liberal egalitarian strain subscribe to all three claims. The second claim, in either one of both its forms, seems hard to deny in the face of surmounting evidence showing the extent to which our success in life (or lack thereof) is correlated with our social and economic status, and with our natural endowments. The third claim is merely a logical conclusion that inevitably follows from the validity of the first and the second claims. Hence if the first claim were as well grounded as the other two, any debate about desert in political philosophy would be futile. Yet, as we have discussed, the brute luck constraint is not a straightforward contention.

Liberal egalitarians have endorsed the brute luck constraint based on the pre-assumed support of it by every person who seriously considers the issue. In this sense, we have seen from Rawls' quotes that he assumed this view to be one of the fixed points of our considered judgments. Freiman & Nichols have showed that such is not always the case. They provided us with an experiment that suggested that the layperson endorsement of the brute luck constraint is dependent on the whether the moral example is described abstractly or concretely. The folk would side with philosophers if the case was described abstractly, and diverge from them if the case were described concretely. Freiman & Nichols tentatively explain this difference by saying that

> Perhaps this [difference] is due to a tendency among laypersons to consider moral questions in concrete cases, whereas philosophers are more likely to directly appraise abstract moral principles. A methodological difference may underwrite the apparent moral difference. (2011, p. 133)

It remains a fact that there is a "*moral divide* between lay persons' and political philosophers' attitudes toward desert" (Freiman & Nichols, 2011, p. 133). On the on hand, the folk regard desert as one of the main principles of distributive justice; on the other hand, political philosophers currently present a general tendency to deny desert any role in their theories. Nonetheless, if the adjustments I made on their abstract case are correct, their results regarding folk intuitions about the brute luck constraint do not replicate. As I have reported, I found no significant difference between the abstract and the concrete scenarios and, remarkably, no shift in the participants' judgments of desert.[55] Regarding natural luck (the only kind of luck for which Freiman & Nichols tested), my findings suggest that the folk do *not* endorse the brute luck constraint neither in the abstract nor in the concrete positive scenarios. If such indeed is the case, the philosopher is alone in supporting this constraint.

Before I set to discuss my findings regarding folk intuitions about desert and *social* luck, I would first like to examine my results concerning natural luck in the light of Freiman & Nichols' account of their own findings. They claim to have explained away the aforementioned *moral divide*, showing that it results from distinct reasoning strategies: while the layperson generally thinks in a concrete manner, the philosopher is trained to think in abstract terms. Yet this explanation vanishes in the face of my new findings. If there is no significant difference between the participants' judgments about desert under abstract and concrete scenarios, and if their judgments consistently differ from the philosopher's, what could be the rationale underlying this difference?

I have a tentative answer to the above question. My hypothesis shares a common feature with Freiman & Nichols account of their results, namely, that it is also related to a difference in the way that the general population makes moral judgments compared to the way in which philosophers make these judgments. The latter are trained to think in terms of rational, logical, and abstract principles. When

[55] Freiman & Nichols findings reveal that the participants' desert judgments shift from 'not deserving' to 'deserving' in the abstract and the concrete scenarios, respectively.

faced with the brute luck constraint, the philosopher immediately analyses if it is possible to establish a proper link – namely, responsibility – between the desert basis of X and the agent who performed the action that generated X and, if such is not possible, the philosopher asserts on rational and logical grounds that the agent is not deserving of X. The common person, on the other hand, does not undergo a rational analysis of the case in order to formulate his or her moral assessment; the common person usually relies in his or her intuitions, which are more often than not affectively charged. People either feel a sentiment of approbation towards the action and its consequences, hence judging the actor as deserving; or the other way around, thus judging the actor as not deserving. Now the issue to be resolved is which is the *right* way of assessing moral judgments–and this is an issue to be discussed in future works.

The second noteworthy finding I want to further investigate is the difference observed in the participant's judgments about fairness and desert when the cases involved not natural luck, but social luck. As already mentioned, while they thought that the influence of natural (genetic) luck did not nullify desert, such was not their opinion about the influence of social luck – participants took differences in social luck to undermine claims of deservingness. If one endorses the sentimentalist account exposed in the above paragraph, this difference is not surprising. While one admires a person's natural talents, no one feels sentiments of approbation towards a person's higher economic status, for instance. This is especially so in the negative scenarios, where the cases described individuals who lacked social luck.

In this context, and in light of these new findings, I believe I am better equipped to respond to Olsaretti's claim that the main weakness one can attach to a contribution theory of desert is that it makes "desert depend solely on the outcome produced and on the fact that the agent brought about that outcome. This goes against the conviction that, for a distribution of differential rewards to be justified by desert, people must first have had a fair opportunity to acquire differential deserts" (Olsaretti, Liberty, Desert, and the Market, p. 72). Freiman & Nichols argued that the layperson does not always endorse the conviction that differential deserts must ultimately be grounded in fair opportunities, which would indeed generate a problem for contribution theorists – as emphasized by Olsaretti. Yet their results were misguided given the problems that I corrected for in their abstract case. Additionally, they did

not distinguish between different types of brute luck, investigating solely the intuitions about natural luck.

Hence, in the light of my findings, I have better resources to undermine Olsaretti's claim. The attribution of desert depends 'solely on the outcome produced and on the fact that the agent brought about that outcome' *only* in the case of natural luck. As a consequence, the contribution theorist does *not* go 'against the conviction that, for a distribution of differential rewards to be justified by desert, people must first have had a fair opportunity to acquire differential deserts.' That is, one can endorse a contribution principle – given the fact that natural luck does not invalidate desert, while at the same time being in favor of a principle of equality of opportunity – given that social luck, on the other hand, does undermine claims of desert.

In this context, Freiman & Nichols question the abovementioned conviction that differential deserts must be grounded in fair opportunities. They claim that such conviction prevails among laypersons only under certain conditions, but not under other conditions. Yet they did not investigate people's intuitions on social luck, so they were not able to realize that this conviction is prevalent among the general population for the cases of social inequalities – but not for natural inequalities.

These are mere tentative hypotheses for a philosophical problem that still requires much more empirical and philosophical work. Yet my hope is to have contributed to the political philosophical debate about desert, illuminating specific features of the folk conception of desert and providing new insights for the establishment of the proper role of desert in principles of distributive justice.

Appendix

IRB Approved Project: **Folk Intuitions about Justice, Responsibility, Desert, and Related Concepts**

Expected Dates: December 2013 through December 2014

Background

Contemporary Political philosophers have developed several theories of distributive justice since Rawls's seminal book *A Theory of Justice* (1971). These theories vary across the many dimensions that involve distributive principles, such as: (i) what is relevant – income, wealth, opportunities, jobs, welfare, utility, etc.; (ii) the nature of the recipients – individuals, groups, classes, etc.; and (iii) on what basis should the distribution be made – equality, maximization, according to individual characteristics, according to free transactions, etc.

Amongst these diverse approaches to distributive justice, the main contemporary theories can be divided into two broad categories: (i) Liberal Egalitarian; and (ii) Libertarian. On the one hand, libertarians are solely concerned with the protection of individual rights, firstly envisaged by Locke as the natural rights to life, liberty and property. On the other hand, liberal egalitarians include all political philosophers whose theories embrace the intrinsic value of autonomy and the consequent relevance of individual liberties, while at the same time acknowledging the value of equality.

Following the work of Rawls, liberal egalitarians made claims of responsibility – and consequently, desert – practically disappear from the justice scene. They argue that most – if not all – of our income and wealth comes from brute luck, and that this is sufficient to show that desert should play no role in determining the distribution of income amongst individuals. In other words, they argue that claims of brute luck are

sufficient to nullify claims of desert. Yet we have reason to doubt that this view is shared by the folk. Moreover, we have reason to doubt that philosophers themselves are entitled to this view. Regarding the former doubt, there is an extensive body of empirical research that shows that claims about desert and responsibility constitute an important part of the folk's concept of distributive justice (Miller, Principles of Social Justice, 2003, Chapter Four); regarding the latter doubt, political philosophers such as David Miller, David Schmidtz, and George Sher have begun to ask the question: do claims of brute luck really nullify claims of desert?

Following Hume, these philosophers appeal to commonsense morality's indifference to the conditions under which desert bases are acquired. Yet if the embracement of desert should rest on commonsense morality, it is imperative to confirm if such is indeed the folk's view. Despite the fair amount of evidence collected by a range of different social scientists on the folk's concept of justice, there is not enough evidence on the nuances of the concept of desert.

As a result, several unanswered empirical questions remain. For instance, do the folk actually believe that brute luck does not nullify claims of desert (as the aforementioned researchers have suggested)? Are there differences in this belief according to different kinds of desert basis – effort, artistic talent, athletic talent, etc.? Are there differences in desert beliefs according to the kind of desert; for instance, economic or moral appraisal?

In an effort to contribute to this research program at the cross roads of political philosophy and political psychology, Freiman & Nichols (2010) designed an experiment to shed light on the following conflict: the tendencies observed among the folk to at the same time "judge individuals' deserts in terms of their performance alone and to restrict such judgments to those products within their control" (Freiman & Nichols, 2010, p.2). Their idea is that this conflict rests on the established asymmetry between judgments made either under abstract or under concrete conditions, and their hypothesis is that "subjects presented with a purely abstract question about desert would be more likely to give responses conforming to the brute luck constraint than subjects presented with a concrete case about a particular individual" (Freiman & Nichols, 2010, p.2). While their findings appear to support

their prediction, there are some issues with their experimental design that I seek to investigate (and avoid) with my present research.

The Present Research

The goal of the present research is twofold: (i) to improve upon the experimental design used by Freiman & Nichols (2010); and (ii) to provide additional data on the nuances of the folk's concept of desert.

The first goal rests on the premise that findings reported in Freiman & Nichols (2010) were driven by a misformulation of the abstract scenario. They only used one abstract case, which involved agents who benefit from genetic advantages. However, Friedman and Nichols did not distinguish between different types of genetic advantages. Moreover, they did not specify which kind of genetic advantage was conducive to the individual's higher level of income. This under specification failed to control for alternative interpretations of their case: participants were unwillingly invited to fill in the details not explicit in the case. Therefore, I am going to design new abstract cases that address the misformulation, so as to test if their hypothesis still holds under the revised experimental design.

The second and related goal is to explore some features of the concept of desert that are ignored in their work. Friedman and Nichols used very few scenarios: only one abstract and two concrete. As a result of this limited number of cases they were unable to explore a wide range of people's intuitions about desert. For instance, they were unable to capture features such as (i) how the basis of desert was generated – was it a result of natural luck or of social luck?; and (ii) what is to be deserved, i.e., the nature of the reward – should it be income or moral appraisal?. Hence my aim is to build on their experiment by providing new scenarios that explore these features.

Experimental Design

Adult subjects will be recruited via: (a) Qualtrics' fee-based panelist service, and (b) Amazon.com's user-fee based mTurk service. In order to participate, subjects must be at least 18 years old. There are no special conditions present in the subject population and no contact information will be requested of subjects.

For the first few rounds of studies, subjects will receive cases about desert, along with some of the following psychometric tools:

- Free Will Inventory (FWI: Nadelhoffer et al., in preparation)
- Just World Scale (JWS: Rubin et al., 1975)
- Social Dominance Orientation Scale (SDO: Pratto et al., 1994)
- Economic System Justification Scale (ESJ: Jost & Thompson, 2000)
- Fair Market Ideology Scale (FMI: Jost et al., 2003)
- Right Wing Authoritarianism (RWA: Altemeyer, 1996)
- The Psychological Entitlement Scale (PES: Campbell et al., 2004)
- The Narcissistic Personality Inventory (NPI: Emmons, 1984)

The abstract cases consist in variations of the following:

- (Intellectual) Suppose that some people make more money than others solely because they genetically have above average intellectual capacities. Do they deserve this extra money? Is it fair that they get this extra money only because they have above average intellectual capacities solely due to genetic advantages?

- (Athletic) Suppose that some people make more money than others solely because they genetically have above average athletic capacities. Do they deserve this extra money? Is it fair that they get this extra money only because they have above average athletic capacities solely due to genetic advantages?

The concrete cases consist in variations of the following:

- (Intellectual) Suppose that John and Paul both want to be software programmers. They both study equally hard. Although math is the greatest talent of both John and Jim, John is always able to naturally come up with more efficient solutions for software programs than Jim because of differences in their genetics. Solely as a result of this genetic advantage, John's programing is much better than Jim's. As a result, John gets a better job and makes more money than Jim. Does John deserve this extra money? Is it fair that John gets this extra money only because he has genetic advantages?

- (Athletic) Suppose that John and Paul both want to be basketball players. They both train equally hard. Although basketball is the greatest talent of both John and Jim, John is always able to naturally come up with better plays than Jim because of differences in their genetics. Solely as a result of this genetic advantage, John's playing is much better than Jim's. As a result, John gets a better position in a better team and makes more money than Jim. Does John deserve this extra money? Is it fair that John gets this extra money only because he has genetic advantages?

I also plan to collect the following demographic information:

- Political views
- Gender
- Age
- Ethnicity
- Education Level
- Religious Affiliation
- Religiosity

Recruitment

The subjects recruited through mTurk and Qualtrics will find the study listed among several other studies from which they can choose.

Compensation

Subjects will receive monetary compensation for their participation. The subjects will be informed of the specific monetary amount before they participate.

Risks

Subjects will not be exposed to any physical risk by participating in these studies.
Subjects' names will not be linked to their data (online surveys do not ask subjects for their names). In addition, participation in this study will NOT result in subjects being exposed to criminal/civil legal action, loss of job/employability, or mandatory or voluntary reporting to an outside agency.
Electronic data will be stored on either the PI's laptop or his desktop computer, which are both password protected. Electronic data will also be stored in the Qualtrics survey software, which is web-based. Qualtrics is a secure site, as access to surveys and data are password protected. That is, all data is accessed only by the owner of the survey (the PI) who must provide a user id and a password. All pieces of data are keyed to that owner identification and cannot be accessed by anyone else. In addition, Qualtrics provides the following security measures at their facilities: 24-hour magnetic card key access with secondary biometric authentication, 24-hour onsite security personnel, hardware is housed within interior of building with no direct exterior access, 24-hour on-site staffed Network Operations Center, digital security cameras and intercom system, and power delivery infrastructure, generator, diesel fuel and telecommunications infrastructure maintained in secured underground concrete vault. IP addresses are collected in order to monitor redundant data, but names, addresses or any other contact information about subjects is not recorded and cannot be asked in a survey. In order to be included in a Qualtrics survey panel, the individual must be 18 years old or older.

In addition, the PI's laptop and desktop are only connected to protected internet servers and password protected internet connections. The PI's laptop and desktop are password protected and are only used by the PI. Any hard copies of data will be kept in a locked file cabinet in the PI's office.

Subjects will be provided with monetary compensation for their participation. They will not be told the specific purpose of the study prior to participation. Subjects will be recruited from the general population, and will NOT be the PI's students, staff, or friends. Subjects will learn of the general purpose of the study and its procedures prior to participation, and they will receive information about the specific purpose of the study in the debriefing.
There are no additional risks.

Consent

Upon choosing to click on the link to the study, subjects will be presented with a general description of the study (general purpose and procedures), expected time duration, amount of compensation for their participation, a statement that their responses will be anonymous and confidential, and a statement that their participation is voluntary and that they can terminate their participation at any time. After reviewing this information, they will choose whether to participate in the study.
The present research presents no more than minimal risk and involves procedures that do not require written consent when they are performed outside of a research setting (i.e., reading descriptions of agent's who decide to take performance enhancing drugs). As this research will be conducted via the internet, subjects will not be able to provide their signature on a consent form.
The present research presents no more than minimal risk to subjects (as discussed in the Risk section), and subjects will still be provided with a general purpose of the study and an explanation of the procedures (thus not adversely affecting the rights and welfare of the subjects). Subjects will be provided with a debriefing statement (see Attachment). In addition to detailing the specific purpose of the study, the debriefing statement will also include the contact information for the PI (who subjects will be

told to contact if they have any further questions about the research) and the Human Subjects Committee at the College of Charleston.

Cost-Benefit Analysis

There are very few possible risks to these studies, and they are greatly outweighed by the benefits this research will have in adding to the understanding of our beliefs about responsibility and desert, and its implications for possible solutions to the pressing issue of distributive justice.

Epílogo

Meu principal objetivo com este trabalho foi defender um papel mais substancial para todos os tipos de evidência empírica sobre nosso comportamento moral no desenvolvimento de teorias de justiça. Eu argumento que, apesar de John Rawls, principal filósofo político do século XX, incorporar alguns tipos de evidência empírica em sua teorização, existe ainda uma necessidade premente de atribuição de um papel mais significativo para as ciências empíricas na filosofia política contemporânea. Nesse intuito, eu apresentei quatro artigos relativamente independentes, mas conectados por um objetivo comum: a defesa de uma filosofia política empiricamente informada. Nesta última seção eu apresento uma visão geral desses quatro artigos, respectivamente, resumindo as lições aprendidas durante o processo de sua conclusão.

No primeiro artigo foram abordados os principais argumentos contrários ao uso de evidências empíricas por filósofos políticos. Nesse contexto, foram apresentadas as duas principais críticas a uma filosofia política empiricamente informada e discutidos, consequentemente, os argumentos suficientes para a rejeição desses criticismos. Por último, e à luz das discussões anteriores, busquei especificar a maneira através da qual os filósofos políticos devem colaborar com os cientistas empíricos.

Concluo que estamos, nesse momento, vivendo sob um novo paradigma, no qual caminhamos na direção de um melhor entendimento de como o nosso cérebro funciona. Como resultado, estamos nos tornando cada vez mais capazes de desenvolver teorias filosóficas políticas para instituições e pessoas *reais*. Nesse sentido, as ciências empíricas proporcionam um elenco de dados relevantes sobre crenças e comportamentos humanos que podem informar de forma profícua aos filósofos políticos no desenvolvimento de suas teorias.

Tendo em vista os argumentos desenvolvidos neste primeiro artigo, defendo que a forma adequada de colaboração entre as ciências empíricas e as teorias normativas é chamada forma de colaboração *simbiótica*. O tipo simbiótico de colaboração sustenta uma relação entre a ética e as ciências empíricas na qual uma complementa a outra em suas limitações. Ou seja, a abordagem simbiótica implica em

uma relação pragmática e colaborativa entre teoria normativa e pesquisa empírica, em que os núcleos de cada abordagem permanecem essencialmente separados.

A partir da abordagem simbiótica, os filósofos políticos não mais podem negligenciar todas as informações empíricas relevantes, tanto das ciências naturais quanto das sociais, no desenvolvimento de suas teorias. Afinal, em última análise, as teorias normativas dizem respeito ao comportamento *real* de instituições e de seres humanos *reais*, no mundo *real*, e não ao *suposto* comportamento de instituições e de seres humanos *idealizados*, em mundos *hipotéticos*.

No segundo artigo, após ter estabelecido a forma adequada de colaboração entre os filósofos políticos e os cientistas empíricos, eu discuto quais tipos de evidências empíricas são relevantes para os filósofos políticos interessados em teorias de justiça distributiva. Nesse intuito, eu considero uma variedade de diferentes áreas de pesquisa que até agora não foram devidamente reconhecidas pelos filósofos políticos e forneço uma extensa revisão sobre a literatura empírica sobre intuições humanas, crenças e comportamentos relacionados com os conceitos de justiça e equidade. Esta revisão incluiu algumas das pesquisas mais significativas envolvendo estes conceitos nas áreas de primatologia, biologia evolutiva, economia experimental, psicologia moral, psicologia política e social, e neurociência. A ideia foi, mais uma vez, defender o valor do trabalho interdisciplinar em filosofia política, uma área que é por natureza multidisciplinar e deve, portanto, ser tratada como tal. Além disso, eu busquei tornar todos estes programas de pesquisa e alguns de seus resultados mais interessantes facilmente disponíveis para os filósofos políticos, de forma a incentivar o desenvolvimento de uma filosofia política empiricamente informada.

Os resultados empíricos apresentados no segundo artigo trouxeram à luz uma série de implicações interessantes para a filosofia política – sem mencionar uma variedade de possíveis caminhos para futuras pesquisas. De início, muitos dos resultados revelados pelas ciências naturais dizem respeito à origem de nossos sistemas morais. Nesse sentido, eles são capazes de esclarecer como regras morais surgiram em sociedades primatas e humanas, ajudando-nos, assim, a entender melhor o que é a moral e como ela evoluiu.

Por exemplo, os biólogos evolucionistas alegam que as nossas regras morais surgiram como uma solução para um problema de jogo cooperativo entre indivíduos auto-interessados. Dentro desse paradigma, biólogos evolucionistas, psicólogos e até mesmo filósofos, foram capazes de mostrar como o comportamento altruísta surgiu

através do uso de ferramentas fornecidas pelo modelo dinâmico de aprendizagem social. Há uma notável semelhança entre este modo de proceder no estudo das normas morais e a abordagem de Hume, retratada em seu *Tratado da Natureza Humana*. Hume viveu em uma época na qual ainda não haviam surgido esses campos de investigação científica; no entanto, a história que ele conta sobre a origem da justiça tem notável semelhança com os jogos evolutivos e as modernas explicações teóricas sobre a moralidade. Isto revela a importância, para os filósofos políticos, de uma melhor compreensão da genealogia das nossas regras morais.

Brian Skyrms (2002) apontou que, assim como os teóricos evolutivos, "Hume estava interessado em como realmente chegamos ao contrato que temos agora. Ele acredita que devemos estudar os processos que levam a um estabelecimento gradual de normas e convenções sociais," acrescentando que a "maneira correta de seguir a filosofia social Humeana moderna é via um modelo dinâmico da evolução cultural e aprendizagem social" (p. 272; tradução própria). Mesmo que se possa discordar de Skyrms sobre a maneira correta de se engajar na filosofia de Hume, é inegável que os filósofos políticos devem usufruir do fácil acesso a todas estas evidências empíricas.

As ciências empíricas vêm também esclarecendo a natureza de nossas decisões morais. E aqui, mais uma vez, somos impelidos na direção de um entendimento da nossa moralidade sob uma perspectiva Humeana. Os resultados empíricos de uma série de disciplinas vêm revelando que as nossas decisões morais são mais relacionadas aos nossos sentimentos do que às nossas capacidades puramente racionais. Nesse sentido, Fleck & de Waal (2002) afirmam que

> A pesquisa com primatas sugere implicitamente que uma ênfase no papel das emoções é ao mesmo tempo perspicaz e precisa – em grupos de primatas indivíduos são motivados a responder aos outros com base nas reações emocionais ao comportamento alheio.
> (p. 20; tradução própria)

Esta ideia não corresponde a uma visão exclusivamente emocional da moralidade humana; seria errôneo equiparar emoções morais à ausência de racionalidade. As emoções que estão envolvidas nos nossos julgamentos morais, como afirmam Fleck & de Waal, são muito complexas e exigem o uso de nossas capacidades racionais. Colocando de forma mais precisa,

> Talvez o fato de que a pesquisa com primatas mostre que a moralidade é uma consequência tanto de nossas necessidades e respostas emocionais, quanto de nossa capacidade de avaliar racionalmente alternativas, seja forte o suficiente para justificar o espaço assumido por uma perspectiva mais integrada da moralidade que reconhece sua base biológica e seu componente emocional, bem como o papel da cognição. (Fleck & de Waal, 2002, p. 21; tradução própria)

Há ainda pesquisadores que permaneceram céticos acerca do uso das evidencias da primatologia para ajudar a compreender a moralidade humana. Fleck & de Waal (2002) respondem a essas críticas da seguinte forma:

> Um chimpanzé acariciando e afagando uma vítima de um ataque ou compartilhando a sua comida com um companheiro faminto, demonstra atitudes que são difíceis de distinguir das de uma pessoa que toma uma criança chorando nos braços, ou que faz trabalho voluntário para alimentar os pobres. Descartar tal evidência como um produto da interpretação subjetiva de 'naturalistas romanticamente inspirados' (Williams, 1989, p. 190) ou classificar todo o comportamento animal como baseado no instinto e apenas o comportamento humano como prova de decência moral, é enganosa. (p. 23; tradução própria)

Da psicologia moral e social, eu ressaltei evidências que também apontam na direção de uma compreensão sentimentalista da moralidade. O Modelo Social Intuicionista, desenvolvido por Jonathan Haidt, descreve nossas decisões morais como raramente causadas de forma direta por nossas capacidades racionais. No entanto, como o próprio Haidt ressalta, o seu modelo é de caráter descritivo. Nesse sentido, o Modelo Social Intuicionista explica como julgamentos morais *são* realmente feitos, e não como julgamentos morais *devem* ser feitos. No entanto, ainda que as análises psicológicas da nossa moralidade estejam restritas ao conjunto de estudos descritivos, elas são de elevada relevância para os filósofos. Afinal, se somos naturalmente dotados de emoções e inclinações que influenciam de maneira substancial nosso comportamento moral, precisamos ter a melhor compreensão possível de quais são essas emoções e como elas funcionam, de modo a, no mínimo, sermos capazes de promover aquelas que devem ser cultivadas e inibir aquelas que conduzem a um comportamento imoral.

Haidt (2001), Greene (2001, 2004), Prinz (2007) e Nichols (2002, 2004) já começaram a utilizar os resultados empíricos para questionar a propriedade de uma filosofia moral estritamente racionalista. Haidt define bem esse questionamento na seguinte passagem:

> Agora sabemos (mais uma vez) que a maior parte da nossa cognição ocorre automaticamente e fora da consciência (Bargh & Chartrand, 1999) e que as pessoas muitas vezes não são capazes de nos descrever como elas realmente chegaram a um julgamento (Nisbett & Wilson, 1977). Ora, nós sabemos que o cérebro é um sistema de conexões capaz de avaliar situações complexas rapidamente (Bechtel & Abrahamsen, 1991). Agora sabemos também que as emoções não são tão irracionais (Frank, 1988), que o nosso raciocínio não é tão confiável (Kahneman e Tversky, 1984), e que os animais não são tão amorais (de Waal, 1996) como pensávamos na década de 1970. O nosso tempo pode, portanto, ser o tempo de lançar um outro olhar sobre a perversa tese de Hume, qual seja, de que as emoções e as intuições morais são as responsáveis pela condução dos nossos julgamentos morais. (2001, p. 355; tradução própria)

Além de apontar na direção de uma filosofia política com um viés sentimentalista, as ciências empíricas também podem nos ajudar a esclarecer a lógica por trás de nosso endosso de determinados princípios morais. Alguns filósofos políticos, como Michael Walzer e, mais recentemente, David Miller, já chamaram a atenção para a impossibilidade de se chegar a um sistema único de regras morais. Eles argumentam que a moralidade humana é inerentemente pluralista: fazemos uso de princípios distintos em nossos juízos morais, de acordo com a respectiva esfera de vida em que a decisão está sendo tomada.

Curiosamente, a biologia evolutiva vem fornecendo aos filósofos evidências científicas de que as nossas regras morais são de fato plurais. De acordo com Krebs (2002), a moralidade humana evoluiu de tal maneira que herdamos programas flexíveis responsáveis pela organização de conjuntos de estratégias condicionais. Estas estratégias são de domínio específico, no sentido de que elas regulam tipos distintos de relações sociais, quais sejam, hierárquicas, igualitárias e íntimas. Dessa forma, os filósofos pluralistas podem ter entendido de modo correto a nossa moralidade; ao que parece, nós de fato evoluímos de forma a endossar princípios morais distintos quando colocados em diferentes contextos.

Psicólogos morais e economistas comportamentais nos trazem mais um aprendizado interessante através de suas pesquisas empíricas. Por exemplo, Sinnott-Armstrong (2006) argumenta que os vieses e heurísticas que essas ciências vêm revelando em nossos julgamentos morais acarretam implicações importantes para a epistemologia moral, na medida em que se torna imprescindível ao filósofo o estudo

desses resultados empíricos de tal forma a poder distinguir sob quais circunstâncias as crenças morais podem ser consideradas justificadas. Nesse sentido, Sinnott-Armstrong (2006) defende que, nos casos em que as nossas crenças são baseadas em processos que não são confiáveis, elas não podem ser consideradas como justificadas.

Nesse sentido, as pesquisas realizadas por psicólogos e economistas comportamentais podem nos fornecer subsídios para duvidar da confiabilidade de certos tipos de crenças, formadas sob determinadas circunstâncias. Dessa forma, os filósofos intuicionistas não podem simplesmente negar a relevância dessas pesquisas empíricas para o seu campo de interesse; é necessário saber se os pressupostos empíricos que fundamentam seus pontos de vista normativos são de fato confirmados pela realidade. Não é possível realizar essa confirmação sem conhecer de modo mais aprofundado a nossa psicologia e, especialmente, os processos através dos quais nossas crenças morais são realmente formadas (Sinnott- Armstrong , 2006).

Tendo em vista todas essas evidências e suas implicações, o meu foco no terceiro artigo que compõem essa tese foi defender uma maior participação dos nossos sentimentos morais em teorias de justiça. Surpreende o fato de que os filósofos políticos contemporâneos tenham permanecido afastados do sentimentalismo moral mesmo em face de todas as evidências apontando na direção de uma explicação mais emocional da nossa moralidade. Nesse contexto, vimos que Frazer (2010) argumenta que a atual desconsideração do sentimentalismo moral (e a aceitação do racionalismo moral) por filósofos políticos é uma consequência direta do receio de incorrer em relatos meramente descritivos de nossas regras morais, desprovidos de poder normativo.

Além disso, como já afirmei no primeiro artigo, este receio está também relacionado com a ausência de uma participação mais substancial de evidências empíricas no desenvolvimento de teorias de justiça contemporâneas. Se somarmos essas duas consequências de uma preocupação com a ausência de força normativa em teorias de justiça, somos naturalmente conduzidos a uma tendência racionalista nas teorias político-filosóficas, tendência esta que tende a se perpetuar devido a consequente falta de necessidade de acompanhar, constantemente, os avanços feitos pelas ciências empíricas.

Frazer (2010) também aponta outra grande preocupação entre os filósofos políticos com relação ao sentimentalismo moral, qual seja, a afirmação de Rawls de que a nossa experiência de empatia nos leva a negligenciar a individuação das

pessoas. Se a empatia de fato esmaece a separação entre os indivíduos, segue que o sentimentalismo moral é incompatível com uma teoria liberal da justiça construída em torno de direitos individuais e da inviolabilidade das pessoas. No entanto, como eu mostro no terceiro artigo, nem a preocupação de cair em um relato descritivo da moralidade, nem o problema da individuação das pessoas, representam um perigo real para o sentimentalismo moral.

Em seu recente livro, *O Iluminismo da Simpatia*, Michael Frazer buscou iniciar o difícil trabalho de construir uma visão mais sentimentalista da justiça. Nas suas palavras, ele ressalta que procura "recuperar o sentimentalismo moral como um recurso para enriquecer a ciência política, a filosofia política e a prática política; um recurso que hoje é muitas vezes esquecido devido à influência generalizada de explicações racionalistas" (Frazer, 2010, p. 4). Na mesma linha, eu argumentei no terceiro artigo que uma consideração séria e comprometida das recentes evidências empíricas sobre a moralidade humana conduzem ao reconhecimento do sentimentalismo moral como a teoria moral disponível mais adequada. Além disso, eu apresentei uma discussão das principais fraquezas da teoria, mostrando que elas podem ser corrigidas através dos refinamentos desenvolvidos por Adam Smith.

Finalmente, no quarto e último artigo, eu apresentei um experimento que buscou investigar o papel de princípios de merecimento em teorias de justiça. Neste último artigo, eu argumento que princípios de merecimento devem ter um papel importante em teorias de justiça, mesmo em face das loterias natural e social. Eu apresento resultados que sustentam essa visão, na medida em que apontam que a grande maioria das pessoas acredita que a loteria natural não invalida a ideia de merecimento. Entretanto, a maioria das pessoas também acredita que a loteria social é capaz de invalidar esses princípios. Dessa forma, a adoção de um princípio de merecimento não pode vir sozinha; deve vir acompanhada de um princípio de igualdade de oportunidades capaz de reduzir tanto quanto possível os efeitos da loteria social.

Referências

Alesina, A. & Angeletos, G.M. (2002). "Fairness and Redistribution: US versus Europe," *Harvard Institute of Economic Research Working Papers*, Harvard Institute of Economic Research.

Alesina, A. & Angeletos, G.M. (2005). "Fairness and Redistribution," *American Economic Review*, vol. 95(4): 960-980.

Alesina, A. & Giuliano, P. (2009). "Preferences for Redistribution," *NBER Working Papers 14825*, National Bureau of Economic Research.

Alves, W. & Rossi, P. (1978). "Who should get what? Fairness judgments of the distribution of earnings." *American Journal of Sociology* 84: 541-64.

Amie, Y. & Cowell, F. A. (1999). *Thinking about Inequality*. Cambridge University Press.

Arnhart, L. (1998). *Darwinian Natural Right: The Biological Ethics of Human Nature*. Albany: SUNY Press.

Avramova, Y. R. & Inbar, Y. (2013). "Emotion and Moral Judgment." *WIREs Cognitive Science* 4:169–178.

Alexander, R. D. (1987). *The Biology of Moral Systems*. New York: Aldine de Gruyter.

Baier, A. (1988). "Hume's Account of Social Artifice – Its Origins and Originality." *Ethics* 98, pp. 757-78.

Carens, J. (1981). *Equality, Moral Incentives and the Market.* Chicago: Chicago University Press.

Chamberlin, E. H. (1948). "An Experimental Imperfect Market." *The Journal of Political Economy*, Vol. 56, No. 2: 95-108.

Comer, R., & Laird, J. D. (1975). "Choosing to suffer as a consequence of expecting to suffer: Why do people do it?" *Journal of Personality and Social Psychology*, 32: 92-101.

Cushman, F.A., Young, L. & Hauser, M.D. (2006). "The Role of Reasoning and Intuition in Moral Judgments: Testing three principles of harm." *Psychological Science* 17(12): 1082-1089.

Damasio, A. R., Tranel, D. & Damasio, H. (1990). "Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli." *Behavioral Brain Research* 41: 81–94.

Damasio, A. R. (1994). *Descartes' error: emotion, reason, and the human brain*. New York: Grosset/Putnam.

Davis, D. & Holt, C. (1993). *Experimental Economics*. Princeton University Press.

Dworkin, R. (2000). *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge: Harvard University Press.

Eskine, K., Kacinik, N. & Prinz, J. (2011). "A bad taste in the mouth: Gustatory disgust influences moral judgment." *Psychological Science* 22: 295-299.

Felt, A. & Cokely, E. (2012). "The Philosophical Personality Argument." *Philosophical Studies* 161: 227-246.

Fleck, J. C. & de Waal, F. (2002). "Any Animal Whatever: Darwinian Building Blocks of Morality in Monkeys and Apes." In: Katz, L. (ed.), *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives.* Bowling Green, Ohio: Imprint Academic.

Foot, P. (1978). *The Problem of Abortion and the Doctrine of the Double Effect in Virtues and Vices.* Oxford: Basil Blackwell.

Frazer, M. (2010). *The Enlightenment of Sympathy: Justice and the Moral Sentiments in the Eighteenth Century and Today*. New York: Oxford University Press.

Freiman, C., & Nichols, S. (2011). "Is Desert in the Details?" *Philosophy and Phenomenological Research*, 82 (1): 121-133.

Frohlich, N., Oppenheimer, J. A. & Eavey, C. (1987). "Choices of principles of distributive justice in experimental groups." *American Journal of Political Science* 31: 606-36.

Frohlich, N. & Oppenheimer, J. A. (1992). *Choosing Justice: An Experimental Approach to Ethical Theory*.

Frohlich, N., & Oppenheimer, J. A. (1996). "Experiencing Impartiality to Invoke Fairness in the n-PD: Some Experimental Results." *Public Choice*, 86: 117 - 135.

Gaertner & Schokkaert. (2010). *Empirical Social Choice: Questionnaire-Experimental Studies on Distributive Justice*.

Gazzaniga, M. S., Bogen, J. E., & Sperry, R. W. (1962). "Some functional effects of sectioning the cerebral commissures in man." *Proceedings of the National Academy of Sciences.* USA, 48, pp. 1765-1769.

Gill, M. & Nichols, S. (2008). "Sentimentalist Pluralism: Moral Psychology and Philosophical Ethics." *Philosophical Issues* 18: 143-163.

Goldberg, J. H., Lerner, J. S. & Tetlock, P. E. (1999). "Rage and reason: the psychology of the intuitive prosecutor." *European Journal of Social Psychology*, 29:781–795.

Greene, J.D., Morelli, S.A., Lowenberg, K., Nystrom, L.E. & Cohen, J.D. (2008). "Cognitive load selectively interferes with utilitarian moral judgment." *Cognition* 107: 1144-1154.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M. & Cohen, J.D. (2004). "The neural bases of cognitive conflict and control in moral judgment." *Neuron* 44: 389-400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. & Cohen, J. D. (2001). "An fMRI investigation of emotional engagement in moral judgment." *Science* 293: 2105–2108.

Haidt, J., Koller, S. H. & Dias, M. G. (1993). "Affect, culture, and morality, or is it wrong to eat your dog?" *Journal of Personality and Social Psychology* 65(4): 613-28.

Haidt, J. (2007). "The new synthesis in moral psychology." *Science* 316: 998–1002.

Haidt, J. (2001). "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108: 814–34.

Haidt, J. (2003). "The moral emotions." In R. J. Davidson, K. R. Scherer, H. H. Goldsmith (Eds.), *Handbook of Affective Sciences*. Oxford: Oxford University Press.

Haidt, J. & Hersh, M. (2001). "Sexual morality: the cultures and emotions of conservatives and liberals." *Journal of Applied Social Psychology* 31: 191-221.

Haines, E. L., & Jost, J. T. (2000). "Placating the powerless: Effects of legitimate and illegitimate explanation on affect, memory, and stereotyping." *Social Justice Research 13:* 219-236.

Hardin, G. (1983). “Is Violence Natural?” *Zygon*, 18: 405-13.

Hare, R. M. (1973). “Rawls' Theory of Justice—I” and “Rawls' Theory of Justice—II.” *Philosophical Quarterly* 23: 144-155 and 23: 241-252.

Harrison, J. (1967). “Ethical objectivism.” In P. Edwards (Ed.), *The Encyclopedia of Philosophy*, Vol. 3 & 4: pp. 71-75. New York: Macmillan.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). “Emotional Contagion.” *Current Directions in Psychological Science,* 2: 96-9.

Herne, K., and M. Suojanen. (2004). “The role of information in choices over income distributions.” *Journal of Conflict Resolution* 48, no. 2: 173-93.

Herne, K. & Mard, T. (2008). “Three versions of impartiality: An experimental investigation.” *Working Paper*. University of Turku, Department of Political Science, Turku, Finland.

Hoffman, E., McCabe, K., Shachat, K. & Smith, V. (1994). “Preference, Property Rights and Anonymity in Bargaining Games,” *Games and Economic Behavior* 7: 346-380.

Hume, David. (2000). *A Treatise on Human Nature*. Eds. David Fate Norton & Mary J. Norton. Oxford Philosophical Texts, New York, Oxford University Press.

Hume, David. (1985). *Essays Moral Political and Literary*. Edited by Eugene F. Miller, Indianapolis: Liberty Classics.

Ichikawa, J. (2011). “Experimentalist Pressure Against Traditional Methodology.” *Philosophical Psychology* 25 (5):743 - 765.

Inbar, Y., Pizarro, D. A. & Bloom, P. (2012). “Disgusting smells cause decreased liking of gay men.” *Emotion*, 12: 23-27.

Inbar, Y., Pizarro, D. A. & Bloom, P. (2009). “Conservatives are more easily disgusted than liberals.” *Cognition and Emotion*, 23: 714-725.

Inbar Y., Pizarro, D. A., Iyer, R. & Haidt, J. (2012). “Disgust sensitivity, political conservatism, and voting.” *Journal of Personality and Social Psychology* 3: 537-544.

Joyce, R. (2008). “What Neuroscience Can (and Cannot) Contribute to Metaethics.” In W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 3: The Neuroscience of Morality,* pp. 371-394. Cambridge, MA: MIT Press.

Jost, J.T., Blount, S., Pfeffer, J., & Hunyady, Gy. (2003). "Fair market ideology: Its cognitive-motivational underpinnings." *Research in Organizational Behavior*, 25: 53-91.

Jost, J.T., Glaser, J., Kruglanski, A.W., & Sulloway, F. (2003b). "Exceptions that prove the rule: Using a theory of motivated social cognition to account for ideological incongruities and political anomalies." *Psychological Bulletin*, 129: 383-393.

Jost, J. T., Banaji, B. & Nosek, A. (2004). "A Decade of System Justification Theory: Accumulated Evidence of Conscious and Unconscious Bolstering of the Status Quo." *Political Psychology*, 25(6): 881-919.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press.

Kahneman, D., Knetsch, J. & Thaler, R. (1986). "Fairness as a Constraint on Profit-seeking: Entitlements in the Market." *American Economic Review*, 76(4): 728-41.

_____. (1990). "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy*, 98(6): 1325-48.

_____. (1991). "The Endowment Effect, Loss Aversion, and Status Quo Bias: Anomalies." *Journal of Economic Perspectives*, 5(1): 193-206.

Kahneman, D. & Tversky, A. (1973). "On the Psychology of Prediction." *Psychological Review*, 80(4): 237-51.

_____. (1979). "Prospect Theory: An Analysis of Decisions Under Risk." *Econometrica*, 47(2): 263-91.

Kant, I. (1997). *Groundwork of the Metaphysics of Morals*. Edited by Gregor, M. J. Cambridge University Press.

Knobe, J. (2007). "Experimental Philosophy." *Philosophy Compass* 2 (1):81–92.

Knobe, J. & Nichols, S. (2007). "An Experimental Philosophy Manifesto." In Joshua Knobe & Shaun Nichols (eds.), *Experimental Philosophy*. Oxford University Press. 3--14.

Koenigs, M., Young, L., Adolphs, R., et al. (2007). "Damage to the prefrontal cortex increases utilitarian moral judgements." *Nature* 446: 908-11.

Kolm, C. (1996). *Modern Theories of Justice*. Cambridge: MIT Press.

Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge: Cambridge. University Press.

Krebs, D. (2002). "Evolutionary Games and Morality." In: Katz, L. (ed.), *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives*. Bowling Green, Ohio: Imprint Academic.

Kronman, A.T. (1981). "Talent Pooling." In J.R. Pennock & J.W. Chapman (Eds.), *Human Rights: Nomos XXIII* (New York: New York University Press): 58-79.

Lerner, M. J. & Simmons, C. H. (1966). "The observer's reaction to the 'innocent victim': Compassion or rejection?" *Journal of Personality and Social Psychology* 4: 203-210.

Lerner, M. J. & Miller, D. T. (1978). "Just world research and the attribution process: Looking back and ahead." *Psychological Bulletin* 85: 1030-1051.

Lerner, M. J. (1965). "Evaluation of performance as a function of performer's reward and attractiveness." *Journal of Personality and Social Psychology* 1: 355-360.

Liao, M.S. (2008). "A Defense of Intuitions." *Philosophical Studies* 140 (2):247 - 262.

Ludwig, K. (2010). "Intuitions and Relativity." *Philosophical Psychology* 23 (4):427-445.

Michelbach, P. et al. (2003). "Doing Rawls Justice: an experimental study of income distribution norms." *American Journal of Political Science*, 47, n.3: 523-539.

Miller, D. (1976). *Social Justice*. Oxford: Clarendon Press, 1976.

Miller, D. (2003). *Principles of Social Justice*. Cambridge, MA: Harvard University Press.

Molewijk, B., Stiggelbout, A. M., Otten, W. et al. (2004). "Empirical Data and Moral Theory: A Plea for Integrated Empirical Ethics." *Med Health Care Philos*, 7: 55-69.

Molewijk, B. (2004). "Integrated empirical ethics: In search for clarifying identities." *Med Health Care Philos* 7: 85-87.

Moll, J., Eslinger, P. J. & Oliveira-Souza, R. (2001). "Frontopolar and anterior temporal cortex activation in a moral judgment task: preliminary functional MRI results in normal subjects." *Arq Neuropsiquiatr* 59: 657–664.

Moll, J. R. & Eslinger, P. J. (2003). "Morals and the human brain: A working model." *Neuroreport* 14: 299-305.

Moore. (1903). *Principia Ethica*. Cambridge: Cambridge University Press; revised edition with "Preface to the second edition" and other papers, T. Baldwin (ed.), Cambridge: Cambridge University Press, 1993.

Nadelhoffer, T. & Nahmias, E. (2007). "The Past and Future of Experimental Philosophy." *Philosophical Explorations* 10 (2):123 – 149.

Nagel, T. (1986). *The View from Nowhere*. New York: Oxford University Press.

Nichols, S. (2002). "Norms with feeling: towards a psychological account of moral judgment." *Cognition*, 84: 221-236.

Nichols, S. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment.* New York: Oxford University Press.

Nichols, S. (2008). "Moral Rationalism and Empirical Immunity." In W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 3: The Neuroscience of Morality,* pp. 395-408. Cambridge, MA: MIT Press.

Nowak, M. A. & Sigmund, K. (1998). "Evolution of Indirect Reciprocity by Image Scoring." *Nature* 393: 573-7.

Olsaretti, S. – ed. (2003). *Desert and Justice*. Oxford: Oxford University Press.

Prinz, J. (2006). "The emotional basis of moral judgments." *Philosophical Explorations* 9: 29–43.

Rawls, J. (1955). " Two Concepts of Rules." *The Philosophical Review* Vol. 64, No. 1, pp. 3-32

Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Rawls, J. (2001). *Justice as Fairness: A Restatement.* Cambridge, MA: Harvard University Press.

Roemer, J. (1998). *Equality of Opportunity*. Cambridge, MA: Harvard University Press.

Rochat, P. et al. (2009). "Fairness in distributive justice by 3- and 5-year-olds across seven cultures." *Journal of Cross Cultural Psychology* 40: 416-442.

Sadurski, W. (2007). "Arbitrariness of Social and Natural Differences: Luck, Lottery, and Equality." *EUI Working Papers*. Online at: http://ssrn.com/abstract=987251

Sanfey, Rilling, Aronson, Nystrom & Cohen. (2003). "The neural basis of economic decision-making in the Ultimatum Game." *Science* 13; 300(5626):1755-8.

Scheffler, S. (1992). "Responsibility, Reactive Attitudes, and Liberalism in Philosophy and Politics." *Philosophy and Public Affairs* 21(4): 299-323, 301.

Schleidgen et al. (2010). Mission: Impossible? On Empirical-Normative Collaboration in Ethical Reasoning." *Ethic Theory and Moral Practice*, 13: 59-71.

Schnall, S., Haidt, J., Clore, G. L. & Jordan, A. H. (2008). "Disgust as embodied moral judgment." *Pers Soc Psychol Bull*, 34: 1096-1109.

Scott, J. P. (1971). *Internalization of Social Norms: A Sociological Theory of Moral Commitment*. Englewood Cliffs, NJ: Prentice Hall.

Scott, J. P., Matland, D., Michelbach, P., Bornstein, B. (2001). "Just Deserts: An Experimental Study of Distributive Justice Norms." *American Journal of Political Science*, Vol. 45, No. 4 (October), pp. 749–767.

Sen, A. (2009). *The Idea of Justice*. The Belknap Press of Harvard University Press: Cambridge, Massachusetts.

Shweder, R. A. (1990). "In Defense of Moral Realism: Reply to Gabennesch." *Child Development,* 61: 2060-2067.

Sinnott-Armstrong, W. (2005). "Moral Intuitionism Meets Empirical Psychology."

Sinnott-Armstrong, W., Young, L., & Cushman, F. A. (2010). "Moral Intuitions as Heuristics." In J. Doris et al. (Eds.), *The Oxford Handbook of Moral Psychology*. Oxford University Press.

Skyrms, B. (2002). "Game Theory, Rationality and Evolution of the Social Contract." In: Katz, L. (ed.), *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives.* Bowling Green, Ohio: Imprint Academic.

Smith, Adam. (1984). *The Theory of Moral Sentiments*. Ed. A. L. Macfie & D.D Raphael, Indianapolis, Liberty Fund.

Sober, E. & Wilson, D. S. (2002). "Summary of Unto Others: The Evolution and Psychology of Unselfish Behavior." In: Katz, L. (ed.), *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives.* Bowling Green, Ohio: Imprint Academic.

Sommers, T. (2010). "Experimental Philosophy and Free Will." *Philosophy Compass* 5 (2):199-212.

Sosa, E. (2007). "Experimental Philosophy and Philosophical Intuition." *Philosophical Studies* 132 (1): 99-107.

Strawson, G. (1998). "Luck Swallows Everything." *Times Literary Supplement*, June 26.

Thaler, R. & Sunstein, C. (2009). *Nudge*. Penguin Books.

Thomson, J. J. (1976). "Killing, Letting Die, and the Trolley Problem." *The Monist* 59: 204-17.

Thomson, J. J. (1985). "The Trolley Problem." *Yale Law Journal* 94: 1395-1415.

Tomkins, S.S. (1963). "Left and right: A basic dimension of ideology and personality." In R.W. White (Ed.), *The Study of Lives* (pp. 388–411). Chicago: Atherton.

Tomkins, S.S. (1965). "Affect and the psychology of knowledge." In S.S. Tomkins & C.E. Izard (Eds.), *Affect, Cognition, and Personality: Empirical Studies* (pp. 72–97). New York: Springer.

Tversky, A. & Kahneman, D. (1974). "Judgment under Uncertainty: Heuristics and Biases." *Science*, September, 185(4157), pp. 1124-31.

_____. (1981). "The Framing of Decisions and the Psychology of Choice." *Science*, January, 211(4481), pp. 453-58.

_____. (1986). "Rational Choice and the Framing of Decisions." *Journal of Business*, October, 59(4), pp. S251-78.

_____. (1991). "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *Quarterly Journal of Economics*, November, 106(4), pp. 1039-61.

_____. (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, October, 5(4), pp. 297–323.

Unger, P. (1996). *Living High and Letting Die: Our Illusion of Innocence*. New York: Oxford University Press.

Valdesolo, P. & DeSteno, D. A. (2006). "Manipulations of emotional context shape moral decision making." *Psychological Science*, 17: 476-477.

de Waal, F. (1982). *Chimpanzee Politics: Power and Sex Among Apes.* London: Jonathon Cape.

de Waal, F. (1997). "The Chimpanzee's Service Economy: Food for Grooming." *Evolution and Human Behavior*, 18: 375-86.

Weaver, G. R. & Trevino, L. K. (1994). "Normative and Empirical Business Ethics: Separation, Marriage of Convenience, or Marriage of Necessity?" *Business Ethics Quarterly* 4: 129-143.

Wheatley, T. & Haidt, J. (2005). "Hypnotically induced disgust makes moral judgments more severe." *Psychological Science* 16: 780-84.

Williams, G. C. (1966). *Adaptation and Natural Selection*. Princeton, NJ: Princeton University Press.

Williamson, T. (2011). "Philosophical Expertise and the Burden of Proof." *Metaphilosophy* 42 (3):215-229.

Wilson, J. (1973). *Introduction to Social Movements*. Basic Books Inc. New York.

Yaari, M. E. & Bar-Hillel, M. (1984). "On dividing justly." *Social Choice and Welfare*, 1: 1-24.