UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS REITOR JOÃO DAVID FERREIRA LIMA
PROGRAMA DE PÓS-GRADUAÇÃO EM SOCIOLOGIA E CIÊNCIA POLÍTICA

Julio Henrique Corrêa Silva

**Unfreedom in Context**

Florianópolis
2023

Julio Henrique Corrêa Silva

**Unfreedom in Context**

<div style="text-align: right">

Tese submetida ao Programa de Pós-Graduação em Sociologia e Ciência Política da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do Grau de Doutor em Sociologia e Ciência Política.

Orientador: Prof. Dr. Ricardo Virgilino da Silva.

</div>

Florianópolis
2023

Julio Henrique Corrêa Silva

**Unfreedom in Context**

O presente trabalho em nível de Doutorado foi avaliado e aprovado, em 30 de novembro de 2023, pela banca examinadora composta pelos seguintes membros:

Prof. Dr. Ricardo Virgilino da Silva
PPGSP/UFSC
Orientador/Presidente

Prof. Dr. Alberto Ribeiro Gonçalves de Barros
FFLCH/USP

Prof. Dr. Jean Gabriel Castro da Costa
PPGSP/UFSC

Prof. Dr. Tiago Daher Padovezi Borges
PPGSP/UFSC

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutor em Sociologia e Ciência Política.

Insira neste espaço a
assinatura digital

_____
Prof. Dr. Luiz Gustavo da Cunha de Souza
Coordenador PPGSP/UFSC

Insira neste espaço a
assinatura digital

_____
Prof. Dr. Ricardo Virgilino da Silva
Orientador

Florianópolis, 2023.

# ACKNOWLEDGMENTS

The most valuable insights are discovered last, but the most valuable insights are the *methods*.

Nietzsche, Antichrist, 13

[...] and you will know the truth, and the truth will set you free.

Jesus, John 8:32

# ABSTRACT

Conceptual disagreements often conceal normative disagreements. In this dissertation, I claim that they should not be conflated, and propose a methodological procedure for distinguishing between them. Part I provides the basis for that claim, and introduces that methodological procedure. Parts II and III illustrate the discussion developed in Part I by providing an analysis of the concept of unfreedom and by applying the proposed methodological procedure to some disagreements over it. More specifically, in Part I, I explain how a method can contribute to first-order normative theorizing; provide a general characterization of normative judgments, outline a model of the content of normative beliefs, and defend it against some potential objections; argue that that model supports a pragmatic account of how the content of normative beliefs is determined, which in turn have important implications for the analysis of disagreements involving normatively used concepts; distinguish between substantive and verbal disagreements; argue that normative disagreements tend to show up as legislatively verbal disagreements; claim that normative disagreements should not be conflated with conceptual disagreements; and propose a methodological procedure to distinguish between normative and verbal disagreements. In Part II, I explicate the concept of unfreedom in terms of limitations on practical activity that limit individual human potential in agency; specify three domains around which such limitations tend to cluster (viz., the formation of the will, the exercise of the will, and option sets); argue that ascriptions of unfreedom are context-sensitive; and argue that context often provides a normative parameter that helps fix a conception of individual human potential in agency, which in turn helps select the limitations underpinning particular ascriptions of unfreedom. In Part III, I apply the methodological procedure introduced in Part I to some verbal disagreements between liberals and republicans in contemporary political theory concerning the concept of unfreedom; argue that the disagreement over whether or not unfreedom requires interference should be diagnosed as merely verbal; argue that the disagreement over whether or not unfreedom only requires interference should be diagnosed as legislatively verbal; and present some normative disagreements that might underlie verbal disagreements over unfreedom.

**Keywords:** Conceptual and normative disagreements. Normative beliefs. Unfreedom. Unfreedom in context. Contemporary political theory.

# RESUMO

Disputas conceituais frequentemente ocultam disputas normativas. Nesta tese, afirmo que elas não devem ser confundidas e proponho um procedimento metodológico para distinguir entre elas. A Parte I oferece a base para aquela afirmação e introduz o mencionado procedimento metodológico. As Partes II e III ilustram a discussão desenvolvida na Parte I por meio de uma análise do conceito de *unfreedom* e da aplicação do procedimento metodológico proposto a disputas acerca deste conceito. Mais especificamente, na Parte I eu explico como um método pode contribuir para a teoria normativa de primeira ordem; caracterizo, em termos gerais, o que são juízos normativos e esboço um modelo do conteúdo de crenças normativas, que defendo contra potenciais objeções; argumento que esse modelo respalda uma explicação pragmática da determinação do conteúdo de crenças normativas, que por sua vez possui implicações relevantes para a análise de disputas envolvendo conceitos usados normativamente; distingo entre disputas substantivas e verbais; argumento que disputas normativas tendem a aparecer como disputas legislativamente verbais; afirmo que disputas normativas não devem ser confundidas com disputas conceituais; e proponho um procedimento metodológico para distinguir entre disputas normativas e verbais. Na Parte II, eu analiso o conceito de *unfreedom* em termos de limitações à atividade prática que limitam o potencial individual humano na agência; especifico três âmbitos ao redor dos quais essas limitações tendem a coalescer (nomeadamente, os âmbitos da formação da vontade, do exercício da vontade e dos conjuntos de opções); argumento que atribuições de *unfreedom* são sensíveis ao contexto; e argumento que o contexto frequentemente oferece um parâmetro normativo que ajuda a fixar uma concepção de potencial individual humano na agência, que por sua vez ajuda a selecionar as limitações subjacentes a atribuições de *unfreedom*. Na Parte III, eu aplico o procedimento metodológico introduzido na Parte I a algumas disputas entre liberais e republicanos na teoria política contemporânea relativas ao conceito de *unfreedom*; argumento que a disputa sobre se *unfreedom* requer interferência deve ser diagnosticada como meramente verbal; argumento que a disputa sobre se *unfreedom* apenas requer interferência deve ser diagnosticada como legislativamente verbal; e apresento algumas disputas normativas que podem estar na origem de disputas sobre *unfreedom*.

**Palavras-chave:** Disputas conceituais e normativas. Crenças normativas. *Unfreedom*. *Unfreedom* em contexto. Teoria política contemporânea.

## RESUMO EXPANDIDO

**Introdução**

Frequentemente, disputas conceituais refletem disputas normativas. Isso torna a análise de conceitos normativamente usados particularmente sujeita a potenciais confusões. Nesta tese, investiga-se a relação entre aqueles tipos de disputa e propõe-se um procedimento metodológico para distingui-las. A utilidade desse procedimento é testada através de sua aplicação a disputas acerca do conceito de *unfreedom* na teoria política contemporânea. Espera-se que as propostas metodológicas avançadas aqui contribuam para tornar as disputas normativas mais explícitas e, dessa forma, para estimular o progresso filosófico.

**Objetivos**

Esta tese possui os seguintes objetivos: (i) fundamentar a prescrição metodológica de que disputas normativas não devem ser confundidas com disputas conceituais; (ii) propor um procedimento metodológico para distinguir entre esses dois tipos de disputa; (iii) investigar, através da aplicação do procedimento metodológico proposto, se algumas disputas acerca do conceito de *unfreedom* na teoria política contemporânea refletem disputas normativas.

**Metodologia**

A persecução dos objetivos (i) e (ii) ocorre na Parte I, e a persecução do objetivo (iii) ocorre nas Partes II e III. A elucidação da relação entre disputas conceituais e disputas normativas requereu a elaboração de um modelo de crenças normativas que utiliza o ferramental oferecido pela lógica dos mundos possíveis. Por sua vez, a elaboração do procedimento metodológico para distinguir entre aquelas disputas baseou-se em técnicas familiares de distinção entre disputas substantivas e verbais. Por fim, a análise do conceito de *unfreedom*, na Parte II, baseou-se principalmente na teoria coerentista da liberdade de Christine Swanton, enquanto a aplicação, na Parte III, do procedimento metodológico proposto mobilizou a vasta literatura que compõe o debate ocorrido nas últimas duas décadas entre proponentes da concepção puramente negativa de liberdade e proponentes da concepção neorrepublicana de liberdade.

**Resultados e Discussão**

Os resultados deste trabalho podem ser divididos em três categorias. A primeira, metodológica, diz respeito à elucidação da relação entre disputas conceituais e disputas normativas e à elaboração de um procedimento para distingui-las. Mais especificamente, mostrou-se que, das três classes de juízos normativos (avaliativos, justificatórios e prescritivos), duas – as classes dos juízos avaliativos e justificatórios, ou simplesmente crenças normativas – podem ser modeladas com ferramentas corriqueiramente utilizadas na lógica dos mundos possíveis: conjuntos e *possibilia*. Esboçou-se, assim, um modelo segundo o qual o conteúdo de uma crença normativa é idêntico a conjuntos de mundos possíveis relativos a uma hierarquia ordenada de mundos possíveis. Uma das principais vantagens desse modelo é sua generalidade: ele é compatível com todas as posições normativas e metanormativas disponíveis, exceto o não-cognitivismo. Mostrou-se, também, que o modelo respalda uma explicação pragmática, em vez de semântica, sobre a determinação do conteúdo

de crenças normativas. Segundo essa explicação, uma crença "torna-se" normativa apenas no estágio pós-semântico: após o estabelecimento dos valores semânticos, um parâmetro contextual fornecendo hierarquias de possibilidades ajuda a determinar quais mundos possíveis devem ser selecionados. Na sequência, apresentaram-se algumas implicações da conjunção desse modelo e da explicação pragmática que ele respalda para a análise de disputas envolvendo conceitos normativamente usados. Num primeiro momento, distinguiu-se entre disputas substantivas, meramente verbais e legislativamente verbais, as duas últimas sendo variedades de disputas conceituais. Num segundo momento, distinguiu-se entre dois tipos de disputas normativas: (i) disputas sobre se um item realiza ou promove um dado mundo possível hierarquicamente ordenado; e (ii) disputas sobre a maneira correta de especificar uma dada hierarquia de mundos possíveis. Argumentou-se que, embora disputas do tipo (ii) frequentemente assumam a forma de disputas legislativamente verbais, elas devem ser tratadas como disputas substantivas. O corolário dessa discussão é a prescrição que organiza este trabalho, a saber, que disputas normativas não devem ser confundidas com disputas conceituais. Por fim, propôs-se um procedimento metodológico constituído de quatro etapas para distinguir entre disputas conceituais e normativas. Essencialmente, esse procedimento consiste em substituir ou bloquear o uso do termo que potencialmente está na origem da disputa e investigar se ela persiste de forma a resultar numa disputa do tipo (ii) descrita acima – que, recomendou-se, deve ser tornada explícita.

A segunda categoria de resultados, conceitual, diz respeito à elucidação do conceito de unfreedom. Mais especificamente, procurou-se analisar o conceito de *unfreedom* de forma que, tanto quanto possível, a análise resultante não dependesse de crenças normativas. Para tanto, fazendo-se uso da teoria coerentista da liberdade de Christine Swanton, analisou-se o conceito de *unfreedom* em termos de limitações à atividade prática que limitam o potencial individual humano na agência. A atividade prática foi caracterizada como um processo de focalização de desejos no mundo através de ação intencional, e o potencial individual humano foi caracterizado em termos de três dimensões, a saber, capacidade, propensão e florescimento. Argumentou-se que a análise proposta captura o padrão subjacente às atribuições de *unfreedom* que possuem maior relevância. A fim de tornar essa análise mais informativa, especificaram-se três âmbitos ao redor dos quais as mencionadas limitações tendem a coalescer. O primeiro é o âmbito da formação da vontade, cujas principais limitações podem ser subsumidas nas noções de irracionalidade e inautenticidade (que inclui vicariedade e dissociatividade); dessa forma, capturou-se a intuição de que existe uma relação estreita entre *unfreedom* e a capacidade organizadora da vontade. O segundo é o âmbito do exercício da vontade (ou da tradução da vontade em ação), cujas principais limitações podem ser divididas entre aquelas que causam o colapso da vontade e aquelas que causam uma incongruência entre vontade e ação; dessa forma, capturou-se a intuição de que existe uma relação estreita entre *unfreedom* e a capacidade executora da vontade. O terceiro é o âmbito dos conjuntos de opções, cujas principais limitações podem ser subsumidas nas noções de indisponibilidade e inelegibilidade; dessa forma, capturou-se a intuição de que existe uma relação estreita entre *unfreedom* e os objetos da atividade prática. Na sequência, argumentou-se que atribuições de *unfreedom* são sensíveis ao contexto: a mesma sentença empregando *unfreedom* pode expressar proposições diferentes dependendo do contexto de

declaração, que ajuda a reduzir a variedade caleidoscópica de limitações relevantes. Em especial, argumentou-se que o contexto frequentemente oferece um parâmetro normativo que ajuda a fixar uma concepção de potencial individual humano na agência, que por sua vez ajuda a selecionar as limitações subjacentes a atribuições de *unfreedom*.

A terceira e última categoria de resultados, substantiva, diz respeito à elucidação do debate entre liberais e republicanos na teoria política contemporânea. Mais especificamente, aplicou-se o procedimento metodológico introduzido na Parte I à disputa acerca do status da noção de interferência como condição para atribuições de *unfreedom*, disputa central no debate ocorrido nas últimas duas décadas entre proponentes da concepção puramente negativa de liberdade e proponentes da concepção neorrepublicana de liberdade. Argumentou-se que a disputa sobre se *unfreedom* requer interferência deve ser diagnosticada como meramente verbal, principalmente porque as partes do debate implicitamente assumem que limitações a conjuntos de opções são necessárias para atribuições de *unfreedom*. Por outro lado, argumentou-se que a disputa sobre se *unfreedom* apenas requer interferência deve ser diagnosticada como legislativamente verbal, principalmente porque os neorrepublicanos, diferentemente dos liberais, sustentam que apenas um subconjunto de limitações a conjuntos de opções – a saber, aquelas que também limitam a formação da vontade de modo a torná-la vicária – é suficiente para atribuições de *unfreedom*. Finalmente, de forma tentativa, apresentaram-se algumas disputas normativas que podem estar na origem de disputas acerca do conceito de *unfreedom*.

**Considerações Finais**
Este trabalho procurou fundamentar a prescrição metodológica de que disputas normativas não devem ser confundidas com disputas conceituais, propor um procedimento metodológico para distinguir entre elas e ilustrar esse procedimento através de sua aplicação a disputas acerca do conceito de *unfreedom* na teoria política contemporânea. Espera-se que as propostas metodológicas avançadas aqui contribuam para tornar as disputas normativas mais explícitas e, dessa forma, para estimular o progresso filosófico.

**Palavras-chave:** Disputas conceituais e normativas. Crenças normativas. *Unfreedom*. *Unfreedom* em contexto. Teoria política contemporânea.

# CONTENTS

**INTRODUCTION**

Political philosophy is always tempted by its own subject - so much so that, sometimes, it could be more aptly named philosophical politics. Philosophy is mainly about the pursuit of the truth, whereas politics is mainly about the pursuit of victory. Sitting uncomfortably between these two contrasting activities, political philosophy is often pulled in opposite directions.

In this dissertation, I address a recurrent temptation of political philosophy, namely, that of pursuing victory by legislating on the use of language. One can pursue the victory for one's normative views about the world - that is, about what is good, bad, right, wrong, and how the world should and should not be - by legislating on how language should be used, in a way that language comes to reflect those views. If this legislative enterprise succeeds, one's prospects of victory increases considerably: language users will become citizens of the realm constituted by those views. When unconstrained, this temptation can be problematic for two reasons: first, to the extent that the truth about those views remains unexamined, it will involve philosophical cheating; second, to the extent that the examination of those views is obstructed, it will turn those citizens into serfs.

My goal in what follows is to provide some methodological constraints that help canalize this temptation in a way that is more congenial to political philosophy. I do so by advancing the claim that normative disagreements should not be conflated with conceptual disagreements, and by proposing a methodological procedure to distinguish between them. My hope is that the application of this procedure helps make explicit normative disagreements, thereby compelling lesligative incursions on language use to be based on explicit normative reasoning.

The dissertation proceeds in three parts. Part I provides the basis for the claim that normative disagreements should not be conflated with conceptual disagreements, and proposes a methodological procedure for distinguishing between them. More specifically, in Part I, I explain how a method can contribute to first-order normative theorizing; provide a general characterization of normative judgments, outline a model of the content of normative beliefs, and defend it against some potential objections; argue that that model supports a pragmatic account of how the content of normative beliefs is determined, which in turn have important implications for the analysis of disagreements involving normatively used concepts; distinguish between substantive and verbal disagreements; argue that normative

disagreements tend to show up as legislatively verbal disagreements; claim that normative disagreements should not be conflated with conceptual disagreements; and propose a methodological procedure to distinguish between normative and verbal disagreements.

Parts II and III illustrate the discussion developed in Part I by providing an analysis of the concept of unfreedom and by applying the proposed methodological procedure to some disagreements over it. More specifically, in Part II, I explicate the concept of unfreedom in terms of limitations on practical activity that limit individual human potential in agency; specify three domains around which such limitations tend to cluster (viz., the formation of the will, the exercise of the will, and option sets); argue that ascriptions of unfreedom are context-sensitive; and argue that context often provides a normative parameter that helps fix a conception of individual human potential in agency, which in turn helps select the limitations underpinning particular ascriptions of unfreedom.

In Part III, I apply the methodological procedure introduced in Part I to some verbal disagreements between liberals and republicans in contemporary political theory concerning the concept of unfreedom; argue that the disagreement over whether or not unfreedom requires interference should be diagnosed as merely verbal; argue that the disagreement over whether or not unfreedom only requires interference should be diagnosed as legislatively verbal; and present some normative disagreements that might underlie verbal disagreements over unfreedom.

## PART I - NORMATIVE AND VERBAL DISAGREEMENTS

### 0. Introduction.

In this Part, I explain how a method can contribute to first-order normative theorizing; provide a general characterization of normative judgments, outline a model of the content of normative beliefs, and defend it against some potential objections; argue that that model supports a pragmatic account of how the content of normative beliefs is determined, which in turn have important implications for the analysis of disagreements involving normatively used concepts; distinguish between substantive and verbal disagreements; argue that normative disagreements tend to show up as legislatively verbal disagreements; claim that normative disagreements should not be conflated with conceptual disagreements; and propose a methodological procedure to distinguish between normative and verbal disagreements.

## 1. First-Order Normative Theories.

First-order normative theories aim at grounding the systematic production of normative judgments. This is the purpose, for example, of utilitarianism and virtue ethics in normative moral theory: to ground the production of normative moral judgments in terms of utility and virtue, respectively. How can a method contribute to that aim?

The conditions under which a method can contribute to first-order normative theorizing depend on how that aim is specified, which in turn largely depends on one's metanormative position. One's metanormative position is determined on two axes, one concerning the nature of normativity, the other concerning the nature of normative judgments. The former axis corresponds to the conjunction of answers to the following questions: are there any normative properties? If there is at least one normative property, of which kind is it? Answers to the first question yield the overarching division between realism, which answers affirmatively, and non-realism, which answers negatively; whereas answers to the second question yield the division between non-reductive realism (or primitivism) and reductive realism. The latter axis corresponds to answers to the following question: do normative judgments express beliefs representing the world? Answers to this question yield the traditional division between cognitivism, which answers affirmatively, and non-cognitivism, which answers negatively. The combination of views along these two axes determines a metanormative position.

Since there is disagreement over which metanormative position is correct, it is natural to expect the specification of that aim (of grounding the systematic production of normative judgments) to reflect this disagreement. For general examples, since realists believe in the existence of normative properties, they will be inclined to think that first-order normative theories should ground the systematic production of normative judgments in normative properties; in turn, since non-realists do not believe in the existence of normative properties, they will be inclined to think either that first-order normative theories should ground the systematic production of normative judgments in normative judgments themselves, or that first-order normative theories cannot be grounded at all. For more specific examples, non-reductive realists will be inclined to emphasize the sui generis normative properties they believe to be ascribed by normative judgments; reductive realists will be inclined to emphasize the descriptive properties they believe to be ultimately ascribed by normative judgments; error theorists will be inclined to emphasize either that first-order

normative theories cannot be grounded at all, or that there is an alternative explanation for our non-referring normative judgments; non-cognitivists will be inclined to emphasize the non-cognitive attitudes (such as approval, disapproval, emotions, plans) they believe to be expressed in normative judgments; and so on.

These various ways of specifying the aim of first-order normative theories can be understood as reflecting one or more of the following four overall goals: (i) to identify which properties, if any, are normative; (ii) to identify which properties, if any, normative judgments refer to; (iii) to explain why we have normative judgments, if there are no normative properties; (iv) to pressure normative judgments into coherence.

Thus, when I say that first-order normative theories aim at grounding the systematic production of normative judgments, I mean the laxed condition of aiming at at least one of these four goals. And a method will be suited to contribute to first-order normative theorizing iff it contributes to the achievement of at least one of these goals.

But what is a normative judgment?

## 2. Normative Judgments.

There are three classes of normative judgments: evaluative, justificatory, and prescriptive. Intuitively, one can grasp the distinction between them in terms of their primary role: when you evaluate, you are primarily rank-ordering items (actions, objects, states of affairs, or whatever); when you justify, you are primarily giving a reason in favor or against an item; when you prescribe, you are primarily instructing the behavior of an item.

More formally, one can grasp the distinction between them in terms of the following necessary conditions: a judgment is evaluative only if it can be conveyed with a sentence that conceptually entails that something is either good or bad; a judgment is justificatory only if it can be conveyed with a sentence that conceptually entails that something is either right or wrong; and a judgment is prescriptive only if it can be conveyed with a sentence that conceptually entails that something should or should not be the case[1]. Therefore, a judgment is normative only if it can be conveyed with a sentence that conceptually entails that something is good, bad, right, wrong, should be the case, or should not be the case.

These are only necessary but not sufficient conditions for being a normative judgment because a non-normative judgment can be conveyed with a sentence conceptually entailing any of the above normative predicates. Belief reports are a case in point: for example, a historian can report on old normative systems by means of sentences conceptually entailing normative predicates without thereby conveying any normative judgment, restricting his report to the conveying of exclusively descriptive beliefs. In fact, the conveying of normative judgments is a highly context-sensitive matter, and shortly I will present a model that attempts to accommodate this fact.

Although many implicitly recognize the distinction between evaluative, justificatory, and prescriptive judgments, rendering it explicit has at least two important consequences. First, it undermines the generally held assumption that only one metanormative position is required in an account of all normative judgments[2]. It could be the case, for instance, that a

---

[1] This is a reconstruction of Streumer's necessary and sufficient conditions for a judgment to count as normative (Streumer 2017, p. 2). The normative properties he lists (idem, p. 105) can either be reduced to the properties of being good, bad, right, wrong, or be ambiguous, as the one ascribed with the predicate permissible, which fails to distinguish between the purely descriptive property of being permitted by someone (say, a community) and the justificatory property that something is right. Unlike him, I admit in my list attributive uses of the predicates good and bad, such as the attributive use of the predicate good in the sentence 'This is a good car'. Finally, for reasons to be presented in the next section, I will avoid talking about the predicate ought.

[2] This consequence does not vindicate a hybrid metanormative position, according to which the same normative judgment is accounted for in terms both of a belief and a non-cognitive attitude. Rather, it points to the

proper account of evaluative and justificatory judgments requires some version of cognitivism, whereas a proper account of prescriptive judgments requires some version of non-cognitivism.

Second, the distinction might be useful to explain contrasts between subject matters with distinctive patternings in normative judgments. For example, epistemology normatively contrasts with psychology in that the former involves mostly justificatory judgments, whereas the latter involves mostly evaluative and prescriptive judgments; ancient political philosophy normatively contrasts with modern political philosophy in that the former involves mostly evaluative judgments, whereas the latter involves mostly justificatory judgments; and so on. This pattern-tracking utility of the distinction between evaluative, justificatory, and prescriptive judgments seems to indicate that it importantly captures some real-world differences.

That being said, it is also the case that most first-order normative theories ground one class of normative judgments in another. This is exactly what the distinction between consequentialism and deontology in first-order normative theorizing amounts to: in general, consequentialist-oriented normative theories ground justificatory and prescriptive judgements in evaluative judgments (e.g. grounding ascriptions of rightness on ascriptions of goodness defined in terms of consequences), whereas deontically-oriented normative theories ground evaluative and prescriptive judgments in justificatory judgments (e.g. grounding ascriptions of goodness on ascriptions of rightness defined in terms of rationality). It is possible, then, that the proposed distinction between three classes of normative judgments stands or falls on whatever first-order normative theory is ultimately proven to be correct. However, this (remote) possibility has no bearing on the present discussion, which aims to be as non-committal in metanormative and first-order normative issues as possible. Moreover, since the proposed distinction might shed light on actual normative judgments as they are currently employed, it will be maintained in what follows.

## 2.1. A Model Of The Content Of Normative Beliefs.

I will now outline a model of normative judgments that builds upon the preceding characterization. However, instead of characterizing normative judgments in terms of sentences conceptually entailing the satisfaction of normative predicates, the model directly

_____

possibility that distinct metanormative positions (hybrid or not) may be required to account for distinct classes of normative judgments.

accounts for the mental states that constitute the content of normative judgments. More specifically, I will model the content of evaluative and justificatory judgments as representations of how the world might be relative to a hierarchical ordering of ways the world might be. In this regard, the model is particularly suited for cognitivist approaches to evaluative and justificatory judgments[3].

It may be helpful to introduce the model metaphorically. Consider first the case of ordinary beliefs. A good way of thinking about the kind of mental state we find ourselves in when holding ordinary beliefs is by analogizing them to maps. Maps provide simplified representations of a given territory that can be put to a variety of uses, but the content of a map is just the simplified representation it provides. Similarly, ordinary beliefs provide simplified representations of how the world might be that can be put to a variety of uses, but the content of a belief is just the simplified representation it provides. Theorists in the possible worlds tradition like to model these ideas with formal objects defined in terms of sets and possibilia. They thus identify the content of ordinary beliefs with possible worlds, sets of possible worlds, sets of centered worlds, and so on, depending on their particular theoretical aims.

This line of reasoning can be helpfully extended to the case of normative judgments. A good way of thinking about the kind of mental state we find ourselves in when holding normative judgments is by analogizing them to navigation systems. Navigation systems provide simplified representations of a given territory relative to a given destination that can be put to a variety of uses (the most obvious being to determine one's position relative to a destination), but the content delivered by a navigation system is just the simplified relative representation it provides.

The crucial difference is that the territory will be represented relative to a given destination: once a destination is established, maps will be selected according to their relevance to achieving that destination, and perhaps according to other criteria as well. For example, if you are traveling by car, you usually want to take the shortest path, but you may also want to avoid dangerous paths (you want to take a road, and not a straight line crossing

---

[3] Why commit to cognitivism? Because language is a tool that can serve many purposes, one of which is to represent the world. So, even if some normative judgments, say, moral ones, are used to express emotions, attitudes of approval or disapproval, to produce perlocutionary effects, or merely to signal group affiliation, I think the representational underpinnings of such uses are worth investigating. Moreover, the model is intended to apply to all normative domains, and attending to domain-specific peculiarities would probably introduce biases that could militate against that intention.

cliffs and volcanoes), travel by the coastline, try a new highway, and so on. If all goes well, the navigation system will deliver instructions that will get you to the final destination through paths satisfying your criteria.

Similarly, normative judgments provide simplified representations of how the world might be relative to a hierarchical ordering of ways the world might be that can be put to a variety of uses, but the content of a normative judgment is just the simplified relative representation it provides. More precisely, the content of evaluative and justificatory judgments is just these relative representations, whereas the content of prescriptive judgments are more like the instructions resulting from these representations. I would like to model these ideas with the same tools routinely used to model ordinary beliefs, that is, with formal objects defined in terms of sets and possibilia.

More specifically, I propose to model the content of evaluative and justificatory judgments (henceforth normative beliefs) as being identical to sets of possible worlds relative to a hierarchy of possible worlds. Both ordinary and normative beliefs thus carve out logical space in a way that excluding possibilities adds information, but normative beliefs necessarily carve out logical space twice, so to speak: first by hierarchizing ways the world might be, second by selecting ways the world might be relative to the resulting hierarchy. Many ordinary beliefs also carve out logical space in more than one step: when you believe a place to be cold you believe it in relation to a temperature scale establishing a threshold for coldness that may vary according to context. What is special about normative beliefs is that they require possibilities to be hierarchically ordered in order to represent ways the world might be.

According to this model, the difference between evaluative and justificatory beliefs lies in how a hierarchy of possible worlds is specified: evaluative beliefs represent the world relative to hierarchies specified in evaluative terms (e.g. in terms of what is good or bad, noble or base), whereas justificatory beliefs represent the world relative to hierarchies specified in justificatory terms (e.g. in terms of what is right or wrong, legitimate or illegitimate). Where such hierarchies come from (God, natural selection, programming, or what have you), and what is their ontological status are questions beyond the purview of this model. Its sole purpose is to account for the structure and content of normative beliefs.

This story about the content of normative beliefs can also be glossed in terms of properties rather than sets and possibilia: normative beliefs ascribe the role property[4] of either realizing or promoting[5] a given hierarchically ordered way of being. In other words, all normative beliefs ascribe the same role property, but their realizing properties depend on how a hierarchy is specified. In turn, how a hierarchy is specified (e.g. in evaluative or justificatory terms), and what is its nature (e.g. whether it contains sui generis properties, only descriptive properties, or no properties at all), are substantive issues on which the model remains silent. Moreover, the disjunctive character of the role property ascribed by normative beliefs makes room for instrumental normative beliefs as well: something can be considered good, bad, right or wrong because it promotes, i.e. is a means towards, rather than realizes, a hierarchically ordered way of being.

What about prescriptive judgments? I think this model does not apply to prescriptive judgments, because prescriptive judgments do not aim at representing how the world might be, but at representing a trajectory for bringing about a possible world. I will not provide a positive proposal for modeling the content of prescriptive judgments, but I want to comment on why this disanalogy between the contents of evaluative and justificatory beliefs on the one hand and prescriptive judgments on the other may go unnoticed. Three reasons stand out.

The first two are obvious and have already been hinted at. First, the very failure to distinguish between types of normative judgments naturally obscures potential disanalogies between them. Second, prescriptive judgments might seem representational in the same way as evaluative and justificatory beliefs because the latter are required by the former. A navigation system can deliver navigation instructions only after a destination, and perhaps other parameters, have already been established. Similarly, prescriptive judgments require the previous establishment of a hierarchy of possibilities, whether they are defined in evaluative or justificatory terms. But the content of prescriptive judgments does not represent how the world might be relative to a hierarchy of ways the world might be; rather, it represents a trajectory (usually in terms of a behavioral sequence) for bringing about a possible world

---

[4] The distinction between role and realizing properties is borrowed from Jackson 1998. A role property is the theoretical role given by a functional definition; so to speak, it is a placeholder for that which satisfies a given theoretical role. A realizing property is that which satisfies that role. For example, temperature is the role property that is satisfied by the mean kinetic energy of a substance's molecule, which is its realizing property.

[5] I deliberately vaguely characterize the second disjunct as 'promoting a given hierarchically ordered way of being' because its precisification depends on metaphysical and first-order issues upon which a model of the content of normative beliefs can remain silent. For example, a reductive realist may be inclined to understand the 'promoting' relation as (non-deviant) causal relation (incidentally, this would have the interesting consequence that a large swath of normative beliefs would be identical to ordinary empirical beliefs).

established by a given hierarchy of possible worlds. Thus, when someone issues a commandment, an injunction, a principle, and so on, he is prescribing trajectories within the context of normative systems establishing the internal relations between different types of judgment.

Third, and more importantly, prescriptive judgments are often expressed with sentences employing the term ought, which is usually taken to express a justificatory concept. This could suggest that prescriptive judgments are representational in the same way as justificatory judgments, and therefore amenable to being similarly modeled. However, the fact that oughts are usually employed in the context of justificatory normative systems - e.g. of hierarchies of possibilities ordered in terms of duties, obligations, what is permissible, impermissible, appropriate, inappropriate, legitimate, illegitimate, and so on - does not reveal anything special about prescriptive judgments; at most, it shows the prevalence of hierarchies of possibilities ordered in justificatory terms. I therefore think that sentences employing the term ought to express normative judgments call for disambiguation between their justificatory and prescriptive senses.

## 2.2. Advantages Of The Model.

This way of modeling the content of normative beliefs has several advantages. First, it applies to all normative beliefs, thus avoiding domain-specific biases. Any view on the nature of normative judgments must take into account the fact that normative thought and talk is pervasive. For example, despite their obvious differences, the sentences 'democracies are good political systems' and 'cannibalism is wrong' intuitively strike us as expressing normative judgments. Therefore, any view on the nature of normative judgments built upon a model that accounts both for their differences and the underlying similarities in virtue of which they are considered normative has an advantage over views built upon domain-specific models. And the model outlined above does exactly that: it accounts for the difference between normative beliefs in terms of their differing ways of hierarchizing possibilities, and it accounts for the underlying similarity in virtue of which they are considered normative in terms of their identical content structure (i.e. sets of possible worlds relative to a hierarchy of possible worlds) or their ascription of the same role property (of either realizing or promoting a given hierarchy). This is why 'democracies are good political systems' and 'cannibalism is wrong' intuitively strike us as being normative, despite being obviously different: both represent how the world is relative to a hierarchy of ways it might be, but they differ in how

and which possibilities are hierarchized. In sum, the model provides a unified account of the pervasiveness of normative beliefs that also makes room for their differences.

Second, the model explains why normative judgments seem to form a special class of judgments. Unlike ordinary beliefs, normative beliefs necessarily require possible worlds to be hierarchically ordered in order to be conveyed. To see this, compare normative beliefs with another class of beliefs with which they are sometimes conflated, viz. those conveyed by judgments of standard. The sentence 'She did not observe the dress code at the Film Festival' is ordinarily taken as conveying a belief that is relative to a standard, which in this case would be provided by the relevant dress code. Were this standard also used to hierarchize possibilities – say, by ranking adherence to the dress code higher in a hierarchy of ways of dressing –, then that sentence could be taken as pragmatically conveying a normative belief – say, the normative belief that 'it was inappropriate that she did not observe the dress code at the Film Festival'. The reason why normative beliefs are sometimes conflated with judgments of standard is, of course, that the main point of the latter is to guide behavior, which often involves the practices of evaluation, justification, and prescription. However, judgments of standard can be normatively inert without ceasing to be judgments of standard. That is, they do not necessarily require possibilities to be hierarchized in order to be conveyed, whereas normative beliefs do. Therefore, normative beliefs can be said to form a special class of judgments in virtue of this requirement.

Third, the previous point sheds some light on the phenomenon of normative leakage, or why normative beliefs are easily pragmatically conveyed: a sentence like 'the child was saved from drowning' can be easily taken to convey the normative belief that 'it is good that the child was saved from drowning', despite not conceptually entailing it, because hierarchies ranking possible worlds in which children are saved from drowning higher than possible worlds in which they are not, are widely shared.

Fourth, the model also sheds some light on the flipside phenomenon of normative leakage, namely, normative suspension. Just as normative judgments are easily conveyed, they are easily suspended. Extant hierarchical orderings of possible worlds can be bracketed and judged against other hierarchies. This is why, for example, philosophers can intelligibly engage in discussions about the value of values. More mundanely, this is why legal apparatuses and social norms more generally can be non-contradictorily normatively assessed, or why psychopaths can understand moral thought and talk. All that is required for

understanding normative beliefs *qua* normative beliefs is that they ascribe the role property of either realizing or promoting a given hierarchy of possible worlds. To this extent, mastery of normative thought and talk is easy and cheap. The difficult and onerous part, of course, is finding the correct specification (if there is any) of a given hierarchy, the correct way of promoting it, and whether a given item fits the bill. This picture has important consequences for analyses of disagreements involving normatively used concepts, and for the results one can expect from such analyses. I will return to these issues in the next section.

Fifth, the model helps explain the phenomenology that usually accompanies some normative judgments. Some normative judgments are usually accompanied by attitudes of approval, disapproval, commendation, condemnation, and so on, *because* they represent the world relative to a hierarchy of ways the world might be. Since answers to the question 'What to do?' involve hierarchization of possibilities, and since those answers usually involve attitudes of approval, disapproval, commendation, condemnation, and so on, it is natural to expect such attitudes to accompany normative judgments. This is especially the case in the normative domains of religion, morality and politics, where answers to the question 'What to do?' often require a high degree of coordination, which in turn requires a high degree of coordination among hierarchies. The panoply of attitudes usually associated with the issuing of normative judgments helps produce the required coordination, both by signaling which hierarchy is being held and, to a greater extent, by influencing an addressee's hierarchy. Again, the bias introduced by those domains in metanormative theorizing seems to underpin some varieties of noncognitivism, such as emotivism, attitude expressivism, and views emphasizing the illocutionary aspect of normative judgments (an emphasis, it is worth noting, that neglects that the production of perlocutionary effects is often more relevant than the expression of illocutionary forces). However, a brief look at other normative domains shows that the attempt to account for normative judgments in terms of their usually accompanying phenomenology is misguided: sentences like 'your solution to the problem was wrong' or 'this is a good recipe for the cake', despite striking us as conveying normative judgments, can hardly be said to elicit that phenomenology. In sum, not only does the model outlined above apply to cases where the usual phenomenology is absent, but it also helps explain when it is present: hierarchies of possibilities do not suffice but may be necessary for eliciting a particular phenomenology, especially under circumstances requiring coordination among them[6].

---

[6] This is not to say that the phenomenology usually accompanying some normative judgments bear no relationship with these judgments. The study of moral sentiments, for example, has proven useful for gaining

Finally, the model is partially noncommittal in metanormative matters and noncommittal in first-order normative matters. That is, it is not committed to answers to the following questions: Are there normative properties? If there are normative properties, what is their nature? What is the correct way of systematizing normative judgments? These are substantive issues on which the model remains silent. However, it does provide subsidies to answering the following metanormative question: 'What is the nature of normative judgments?'. Since it models the content of evaluative and justificatory beliefs as being identical to sets of possible worlds relative to a hierarchy of possible worlds, it supports a cognitivist answer to this question according to which at least evaluative and justificatory judgments aim at representing the world. As a result, the model leaves almost everything as it is – i.e. all varieties of realism and irrealism, and all varieties of first-order normative theories remain available options –, with the considerable advantage of preventing metaphysical and first-order discussions from being mistaken for discussions about judgments. In other words, by being noncommittal in the sense just described, the model helps locate discussions where they belong by excluding where they do not belong.

## 2.3. Potential Objections To The Model.

I turn now to some potential objections to the model outlined above. You may object that the model is not noncommittal in the sense I said it is, that it does not support cognitivism, or that it entails the falsity of the thesis that there is a is/ought gap.

You may object that the model is not noncommittal in the sense I said it is because you might think it entails the truth of some first-order normative view or the truth of some view on the nature of normative properties.

You might think that the model entails the truth of relativism, because you believe in the following conditional: if the truth of normative beliefs is relative to hierarchies of possible worlds, then relativism is necessarily true. As a first-order normative view, relativism requires the truth of normative beliefs to be relativized either to a given set of conventions or to the preferences of a judgment-maker[7]. The model, however, is silent on the correct way of

---

insight into morally ordered hierarchies. Moreover, emotions and other noncognitive attitudes may themselves be used in the specification of some hierarchies (e.g. when a psychiatrist is evaluating the proportionality of a person's emotional responses to certain events, or when moral theorists emphasize reactions and motivations in their moral theories). The point is merely that there is little reason to believe the relationship of such phenomenology with normative judgments to be a conceptual one.

[7] Of course, there is a sense in which the model is relativistic, viz. in the sense that the content of normative beliefs is relative hierarchical orderings of possibilities. But, thus understood, the model does not involve any answer to whether or not normative beliefs pick out real properties, nor to which properties these would be. In

specifying hierarchies of possible worlds. Indeed, it is compatible with hierarchies being specified in terms of conventions or preferences, but it is also compatible with hierarchies being specified in terms of, say, objective moral order, utility, solutions to cooperation problems, longevity, predictive and retrodictive power, God, and so on. Thus, even if relativism provides the correct specification of hierarchies for certain domains of normative thought and talk, as it plausibly does for some judgments of taste, it might not provide the correct specification of hierarchies for, say, some moral and epistemic normative beliefs. More radically, and more implausibly, the model is compatible with there being only one objective hierarchy grounding the truth of all normative beliefs. In sum, the conditional above is false, because the antecedent does not specify which and how possible worlds are hierarchized, whereas the consequent does.

Alternatively, you might think that the model entails the truth of consequentialism, because you believe in the following conditional: if normative beliefs ascribe the role property of either realizing or promoting a given hierarchically ordered way of being, then consequentialism is necessarily true. As a first-order normative view, consequentialism requires items to be assessed according to their consequences on other items or on ulterior states of affairs. The model, however, is silent on the correct way of specifying hierarchies of ways of being. Indeed, it is compatible with hierarchies being specified in terms of consequences, but it is also compatible with hierarchies being specified in terms of, say, permissibility, duties, appropriateness, reasonableness, rightness, and so on, and realizing a hierarchy specified in such terms would simply amount to having the property of, say, being right. Thus, even if consequentialism provides the correct specification of hierarchies for certain domains of normative thought and talk, as it plausibly does for some normative political and medical beliefs, it might not provide the correct specification of hierarchies for, say, some moral and epistemic normative beliefs. However, if normative beliefs only ascribed the property of *promoting* a hierarchy of ways of being, as opposed to the property of either realizing or promoting a given hierarchically ordered way of being, then consequentialism would be necessarily true, even for hierarchies specified in deontic terms, because it would entail that all normative beliefs are instrumental. In sum, the conditional above is false,

_____

the sense of a first-order normative view, however, relativism is a form of realism: it maintains that normative beliefs do pick out, and are made true by, normative properties, but that these properties are relative either to a set of conventions or to the preferences of a judgment-maker - e.g., in some varieties of hedonism or in some accounts of judgments of taste. Note also that, in ordinary parlance, this latter sense of relativism is often conflated with nihilism, the view according to which there are no normative properties. But the metaphysical difference should not be overlooked by philosophers: relativism is a weak form of realism, whereas nihilism is a form of irrealism.

because the antecedent does not specify which and how possibilities are hierarchized, whereas the consequent does, and because the antecedent does not restrict normative beliefs to instrumental beliefs, whereas the latter does.

Alternatively, you might think that the model entails the truth of reductive realism, because you believe in the following conditional: if normative beliefs ascribe the role property of either realizing or promoting a given hierarchically ordered way of being, then normative realism is true. As a metanormative view on the nature of normative properties, normative realism requires the existence of either sui generis normative properties (nonreductive realism) or normative properties that are identical to descriptive properties (reductive realism). The model, however, is silent on the correct way of specifying hierarchies of ways of being. Indeed, it is compatible with hierarchies being specified in terms of sui generis normative properties or descriptive properties, in which case realizing such hierarchies would amount to realizing such properties, but it is also compatible with hierarchies being specified in terms of noninstantiated properties, say in terms of sin or heaven (understood literally), in which case the model would entail the truth of an error theory of these normative beliefs according to which such beliefs fail to refer. Thus, even though the model descriptively specifies the role property ascribed by normative beliefs, viz. the property of either realizing or promoting a given hierarchically ordered way of being, it remains silent on the nature of its realizing property. In sum, the conditional above is false, because the antecedent does not specify which and how possibilities are hierarchized, whereas the latter does.

I therefore think that the objection against the noncommittal nature of the model fails: I think that the model is compatible with any first-order normative view and with any metanormative view concerning the nature of normative properties.

You may also object that, against what I said, the model does not support cognitivism, because it locates the distinctiveness of normative beliefs in how they carve out logical space, rather than in what is thus carved out. In other words, you may think that, according to the model, normativity is a feature of our beliefs, rather than a way the world might be. But this begs the question as to whether or not hierarchies of possible worlds independently exist, a metaphysical issue on which the model remains silent. Moreover, there is an intuitive difference between expressing nonrepresentational mental states, expressing representations that do not ascribe properties, and expressing representations that ascribe properties. The first is the subject of garden-variety forms of expressivism, such as emotivism

and accounts centered on non-assertoric illocutionary forces. The second is the subject of varieties of expressivism focusing on devices for organizing logical space, e.g. questions or mathematical theories. The third is the subject of garden-variety forms of cognitivism. As a consequence of there being forms of expressivism that appeal to representations, representationalism does not capture what is distinctive of cognitivism. Instead, cognitivism is best characterized as dealing only with that subclass of representations that can be in principle adjudicated by the actual world. Therefore, whether or not the model supports cognitivism hinges on a metaphysical view on whether or not there are hierarchies of possible worlds. If you hold the view that there are such hierarchies, the model straightforwardly supports cognitivism. If you hold the view that there are no such hierarchies, either the model supports a specific strand of cognitivism, viz. an error theory according to which normative beliefs ascribe properties that do not exist, or the model supports a hybrid view according to which normative beliefs are composed both by representations of how the world might be and by representations that cannot be in principle adjudicated by the actual world (i.e. that do not add or exclude information), but only aim at organizing logical space in a specific way (in this case, hierarchically). In other words, pending resolution of these metaphysical issues, the model supports a view that is at least partly cognitivist. I therefore think that the objection that the model does not support cognitivism fails.

Finally, you may also object that the model entails the falsity of the thesis that there is a is/ought gap, because you believe in the following conditional: if sentences using the term ought express justificatory beliefs, then there is no is/ought gap. Indeed, the model entails the truth of this conditional. However, if sentences using the term ought express prescriptive judgments, the is/ought gap not only is preserved as is explained: evaluative and justificatory beliefs represent how the world might be relative to hierarchies of possible worlds, whereas prescriptive judgments do not represent how the world might be at all[8]. Since the thesis that there is a is/ought gap is more naturally understood as taking sentences using the term ought to express prescriptive judgments rather than justificatory beliefs, I think the objection fails.

## 2.4. A Pragmatic Account Of How Normative Beliefs Are Determined.

How is the content of normative beliefs determined? I think the model just presented supports a pragmatic, rather than semantic, account of how the content of normative beliefs is determined. More specifically, I think that it supports an account according to which context

---

[8] See the discussion above on why the model does not seem to apply to prescriptive judgments.

provides a parameter containing a hierarchical ordering of possible worlds relative to which sets of possible worlds are selected, and that this parameter is not semantically required; hence it being pragmatic.

Speakers think and talk against a contextual background of presuppositions that helps determine which beliefs they hold and convey[9]. Hierarchies of possibilities are best understood as being encoded in that contextual background, rather than being encoded in the semantic values of words and sentences. On this account, a belief 'becomes' normative only at a post-semantic stage[10]: after semantic values have been established, a contextual parameter providing hierarchies of possibilities helps determine which possible worlds should be selected. Equivalently, but in terms of properties: it is only after this contextual parameter enters the picture that a belief can ascribe the role property of either realizing or promoting a given hierarchy of possible ways of being, and thereby 'become' normative.

I mentioned earlier that the model sheds some light on the phenomena of normative leakage and suspension: normative beliefs can be easily conveyed because many hierarchies of possibilities are widely shared, and normative beliefs can be easily suspended because hierarchies of possibilities can be intelligibly bracketed. The pragmatic account on hand sheds further light on both phenomena. The reason why the sentence 'the child was saved from drowning' can be easily taken to convey the normative belief that 'it is good that the child was saved from drowning', despite not conceptually entailing it, is that that sentence elicits a hierarchy that is typically encoded in speakers' contextual background, viz. a hierarchy ranking higher possible worlds in which children are saved from drowning. And the reason why the normative belief expressed by the sentence 'saving children from drowning is good' can be easily suspended is that, despite eliciting a hierarchy that is typically encoded in speakers' contextual background, that sentence's semantic value does not depend on that hierarchy, since the latter only enters the picture at a post-semantic stage; therefore, the hierarchy can be bracketed, and the normative belief suspended, without that sentence being rendered unintelligible. As I said earlier, mastery of normative thought and talk is easy and cheap. The difficult and onerous part, of course, is finding the correct specification (if there is

_____

[9] See also Dyke 2003; Väyrynen 2013; Cepollaro & Sojanovic 2016. However, unlike Dyke 2003, I do not take normativity to be semantically encoded in character functions; and, unlike Cepollaro & Stojanovic, I do not take normativity to be semantically encoded in presuppositions. The account provided here is closer to Väyrynen's account (2013; 2014), which arrives at similar ideas by a more empirically oriented linguistic approach.
[10] See Rabern 2017; Yalcin 2022.

any) of a given hierarchy, the correct way of promoting it, and whether a given item fits the bill. In other words, the difficult and onerous part is providing a first-order normative theory.

What about our intuition that some concepts are normative? On the pragmatic account on hand, this intuition can be explained by the fact that, typically, normatively used concepts pragmatically elicit entrenched hierarchies relative to which sets of possibilities are selected. For example, the concepts expressed by the predicates pedophilia, cannibalism, betrayal, tyranny, and torture elicit entrenched hierarchies that rank lower possible worlds containing those features. In other words, the properties ascribed by those predicates are all widely accepted to be bad or wrong, but none requires badness or wrongness to be semantically encoded. Similarly, the concepts expressed by the predicates freedom, courage, wisdom, temperance, and justice elicit entrenched hierarchies that rank higher possible worlds containing those features. In other words, the properties ascribed by those predicates are all widely accepted to be good or right, but none requires goodness or rightness to be semantically encoded. Such entrenched hierarchies help guide conceptual usage, but pragmatically, rather than semantically.

The model presented above, and the pragmatic account that it plausibly supports, have important implications for the analysis of disagreements involving normatively used concepts. I will now argue that, despite often showing up as verbal (conceptual)[11] disagreements, normative disagreements should not be conflated with verbal disagreements; then I will propose a methodological procedure for distinguishing between them.

---

[11] In this dissertation, I use conceptual and verbal disagreements interchangeably.

### 3. Disagreements Involving Normatively Used Concepts.

### 3.1. Substantive Versus Verbal Disagreements.

I begin by distinguishing between substantive and verbal disagreements[12]. When people disagree over a sentence S, they either have a substantive disagreement, a verbal disagreement, or both. Substantive disagreements occur when parties necessarily assert incompatible contents, that is, when they necessarily represent the world in incompatible ways. This is the sort of disagreement we have in mind in most ordinary situations in which one party asserts $P$ and another party asserts $\neg P$, such as when parties disagree over whether or not a certain medication prevents transmission of a given disease. In such cases, parties agree on what is required for a given representation to be true, but disagree on whether an item satisfies such requirements; in principle, their disagreement can be adjudicated by the actual world. In other words, parties to substantive disagreements agree on matters of semantic meaning, but disagree on matters of fact[13].

On the other hand, verbal disagreements occur when parties fail to agree on what is required for a given sentence to be true in the first place, and this as a result of disagreeing on the meaning of at least one term T in S. That is, when people verbally disagree over a sentence S this is due to them disagreeing over how to use at least one term T in S. Among verbal disagreements, one can distinguish those that seem pointless from those that seem to have a point. Call the first kind merely verbal disagreements, and the latter kind legislatively verbal disagreements.

In merely verbal disagreements we usually characterize parties as talking past each other: we know that their disagreement over S would disappear should they agree on the usage of T, and that nothing of importance for how they represent the world hinges on their disagreement. In other words, merely verbal disagreements are devoid of representational significance, since they merely reflect contrasting word usage, and can be predicted to disappear should the conflicting parties agree on the usage of T.

In some verbal disagreements, however, the point is precisely to legislate on how to use a given term T, either because language itself is the subject of primary concern for the parties (e.g. for linguists or philosophers seeking conceptual clarity), or because the term is a proxy for other substantive disputes (e.g. for practical, normative, or legal disputes about

---

[12] The ideas presented in this subsection are borrowed from Chalmers 2011. See also Plunkett & Sundell 2013.
[13] Excluding, of course, facts pertaining to language use.

which cases get covered by such-and-such use of the term). In other words, legislatively verbal disagreements have ulterior significance, since they reflect ulterior substantive disputes (e.g. linguistic, practical, legal, normative), and can be predicted to persist in a distinctive way: the conflicting parties would not be able to agree on the usage of T precisely because their ulterior substantive disputes legislate contrasting roles to T.

How to distinguish between these kinds of disagreement? A simple procedure is to either bar or replace T and see whether the disagreement persists, and how it persists: if the disagreement disappears, there is evidence for diagnosing it as merely verbal; if the disagreement persists in a way that can be in principle solved by the acquisition of further information about the world, there is evidence for diagnosing it as substantive; if the disagreement persists in a way that cannot be in principle solved by the acquisition of further information about the world, there is evidence for diagnosing it as legislatively verbal.

For example, suppose that parties A and B disagree over the following sentence: 'The proposed medical intervention is a vaccine'. That is, parties A and B disagree over whether or not a given medical intervention is a vaccine: party A says it is, and party B says it is not. Suppose further that they are not sure about their respective background information on that medical intervention, nor about their respective usage of the term vaccine. They could either bar or replace the potentially contentious term and see whether their disagreement disappears or persists.

Consider a situation where they bar the use of the term vaccine, and use instead a description of the concept that A has in mind. Party A could then reformulate the original sentence in the following way: 'The proposed medical intervention instructs the immune system to identify, and respond to, a given pathogen'. If their disagreement disappears, this is evidence that they were merely verbally disagreeing: perhaps B uses the term vaccine in a more fine-grained way than A, such that it was concealing B's agreement with the belief that A was trying to convey.

If their disagreement persists, this can be evidence that their disagreement is either substantive or legislatively verbal, depending on whether or not it can be in principle solved by the acquisition of further information about the world. For example, B could retort to the reformulated sentence that the proposed medical intervention does not instruct the immune system to identify, and respond to, a given pathogen. In this case, their disagreement would be substantive: in principle, it can be solved by the acquisition of further information about the

world, namely, information as to whether or not the intervention instructs the immune system in that way.

Alternatively, their original disagreement could persist despite agreement on the reformulated sentence. B could argue that, despite agreeing with the belief that A was trying to convey, that belief should not be expressed with the original sentence, because the term vaccine should be used in a more fine-grained way: it should encompass only those interventions relying on inactivated or weakened forms of a pathogen as a means to instruct the immune system, since this subset of interventions allows for well known patterns of prediction, testing, and so on. In this case, their disagreement would be legislatively verbal: it arises precisely because they disagree over which role should be ascribed to the term vaccine.

Disagreements involving normatively used concepts typically show up as legislatively verbal disagreements. Before presenting a methodological procedure to handle them, it is necessary to get a clearer picture of normative disagreements.

### 3.2. Normative Disagreements.

There are two types of normative disagreement:

(i) disagreement concerning whether or not an item satisfies the role property of either realizing or promoting a given hierarchically ordered possible world;

(ii) disagreement concerning the correct way of specifying a given hierarchy of possible worlds.

The first is straightforwardly substantive in the sense characterized above: it involves the assertion of incompatible contents because people disagree on how the world might be normatively. The second is prone to give rise to verbal disagreements, but I think they should be treated as substantive disagreements concerning metanormative or first-order normative issues. That is, normative disagreements concerning the correct way of specifying a given hierarchy of possible worlds will generally amount either to a dispute about the existence and nature of normative properties, or to a dispute about which first-order normative theory is correct. Pending resolution of such disputes, I think verbal disagreements rooted in (ii) should be treated as substantive disagreements.

To see the difference between (i) and (ii), consider a disagreement over the sentence 'X is a good social institution'. This sentence will generally be used to convey the evaluative

belief that X either realizes or promotes a possible way of being a social institution that ranks high in a hierarchy of ways of being a social institution. On the one hand, type (i) disagreements over it will amount to the assertion of incompatible evaluative beliefs by parties who agree on the specification of a hierarchy of ways of being a social institution. This situation would obtain, for example, in a disagreement between two average utilitarians: presumably, both would hierarchize ways of being a social institution in terms of average utility, and therefore any disagreement would amount to their incompatible beliefs about whether or not X realizes or promotes a possible world ranked in terms of average utility. Just like with any substantive disagreement, provision of additional information about the world should suffice for resolving their dispute.

On the other hand, type (ii) disagreements over that sentence will amount to a verbal disagreement over what counts as a good social institution. This situation would obtain, for example, in a disagreement between an average utilitarian and a Rawlsian: presumably, the former would hierarchize ways of being a social institution in terms of average utility, whereas the latter would hierarchize ways of being a social institution in terms of Rawls' two principles of justice. As a consequence, they would be conveying distinct evaluative beliefs with the same sentence, and analyzing a disagreement between them would first require addressing their verbal disagreement over what counts as a good social institution. However, as is well known, the parties to this verbal disagreement would most likely see themselves as having a substantive disagreement concerning what is the correct first-order normative theory for social institutions, at the center of which would be a dispute about the correct way of specifying a hierarchy of possible ways of being a social institution. Similarly, it is often the case that type (ii) disagreements underlie legislatively verbal disagreements, and that the latter function as proxies for first-order normative or metanormative disputes.

**3.3. Methodological Procedure For Distinguishing Between Normative And Verbal Disagreements.**

In order to distinguish between normative and verbal disagreements, I propose the following methodological procedure.

First, investigate whether parties to the disagreement are engaged in a merely verbal disagreement. This can be done by either barring or replacing the term that is potentially giving rise to the disagreement. If the original disagreement disappears, there is evidence that the parties are engaged in a merely verbal disagreement. If the original disagreement persists,

there is evidence that the parties are engaged either in a substantive or in a legislatively verbal disagreement.

Second, investigate whether the remaining disagreement can in principle be solved by the provision of additional information about the world. If it can, there is evidence that the original disagreement is substantive. If it cannot, there is evidence that the original disagreement is legislatively verbal.

Third, if there is evidence that the original disagreement is legislatively verbal, investigate whether it is rooted in a type (ii) normative disagreement. Again, this can be done by either barring or replacing the term that is potentially giving rise to the disagreement. If the original disagreement bottoms out in a disagreement concerning how parties hierarchize possibilities – which usually takes the form of a disagreement over sentences employing basic normative predicates such as good, bad, right, wrong –, there is evidence that a normative disagreement underlies the original verbal disagreement.

Fourth, if there is evidence that a normative disagreement underlies the original verbal disagreement, make it explicit. In this way, parties can directly address their disagreement over the correct way of specifying a hierarchy of possible worlds.

Of course, this methodological procedure is fallible and in many cases unnecessary – the presence of a normatively used concept often suffices to indicate the presence of an underlying normative disagreement. However, by following these steps one can be more certain that the relevant issues will be unearthed, and thereby increase the likelihood of philosophical progress.

Consider, for example, the following disagreement between readers of Shakespeare's *The Tragedy of Julius Caesar*:

A: Brutus acted courageously.

B: No, he didn't.

A and B know that they possess exactly the same information about Shakespeare's Brutus. Assume that they also know the distinctions and analytical procedures presented above. They thus reasonably suspect that they are verbally disagreeing about the use of the term courageous. Their exchange could proceed as follows:

A: Let's bar the term courageous and see whether our disagreement persists. Do you agree that Brutus accepted the risk of incurring great costs for the sake of a goal that he valued even more greatly, despite fear? More specifically: that he accepted the risk of losing his reputation, family, and life for the sake of preventing Rome from becoming a tyranny, and that he did so despite whatever fear someone as stoic as him could experience?

B: Let me clarify my position. I agree that Brutus accepted those costs for the sake of that goal, despite fear. However, queuing for repeatedly stabbing a trustful, unarmed, and unassisted friend does not involve immediate risks for the perpetrator. At the very least, the action around which the play revolves fails to satisfy the condition of risk acceptance. Therefore, my claim, properly understood, is that Julius' murder cannot count as an instance of Brutus' courage.

A: I appreciate your clarification, but it does not solve our disagreement, for the following reasons. Brutus' actions, including the one you mentioned, didn't happen in the void; they happened in a political context that made those costs and goals salient, even if the undertaking of the action did not involve any immediate risk for him. Julius' killing was a political statement, that is, it was intended to state that some members of the aristocracy would not tolerate the republican regime to degenerate into tyranny, even under someone of Julius' stature. Hence the infamous repeated stabbing: it was necessary to ensure the equal implication of the relevant political players. In short, put in its proper context, that political statement was accompanied by a clear recognition and acceptance of a high risk of deadly retaliation.

B: Before addressing the issue of whether or not Brutus' action satisfies the condition of risk acceptance, there is a subtlety in the play that should be remarked upon. Unlike Roman historiography, the play is less conclusive about the adequacy of the conspirators' assessment of Julius' ambitions: there is some evidence that he is arrogant, but there is also evidence that he rejects being proclaimed autocrat. The only evidence of his tyrannical ambitions is the conspirators' assessment itself. Since this subtlety could further complicate our debate by raising questions about belief adequacy and the relative weight of costs and benefits in ascriptions of courage, I'm happy to put it aside for the moment. You were saying that the issue of risk acceptance should be addressed by taking into account the political context within which Brutus' actions occurred; and that, placed in that context, Julius' killing involved the acceptance of a high risk of deadly retaliation. I agree with all of

this, but I'm not convinced that the almost riskless killing of Julius counts as an instance of Brutus' courage. In fact, I think that Brutus himself shares my opinion.

A: How so?

B: When the conspirators are plotting their attack, they entertain the possibility of murdering some of Julius' supporters as well. Brutus promptly rebuffs the idea, on the grounds that he sees their role as that of sacrificers and purgers, rather than butchers and murderers; and that Julius' sacrifice should be seen as an offering to propitiate the gods, or as an action following from their duty to stave off the spirit of tyranny, rather than as an assault on Julius' body. Brutus' reasoning indicates that he does not see the murder as a particularly risky endeavor; instead, he sees it as an unfortunate necessity.

A: Granted, but you are ignoring that Julius' killing involved inordinate psychic costs for Brutus, which attests to his extraordinary willpower.

B: Should we conclude, then, that we were engaged in a merely verbal disagreement all along? After all, my denial of your statement relied on a conception of courage that emphasizes immediate risk of physical harm, whereas you seem to rely on a conception of courage that admits diachronic risk and psychic costs.

A: Perhaps, but only if you admit that your initial view on Brutus' actions was mistaken.

B: In which way was it mistaken?

A: By neglecting that they instantiate the fact that Brutus was a virtuous citizen and man.

At this point, it is clear that A and B are engaged in a legislatively verbal disagreement, probably as a result of a type (ii) normative disagreement concerning the correct way of specifying a hierarchy of ways of being a man and/or a citizen. Presumably, if they move to their underlying normative disagreement, they will engage in consequentialist and justificatory reasoning. They might also argue about character dispositions, perhaps emphasizing the Platonic requirement that courage be accompanied by other virtues in order to become a virtue; or the Aristotelian notion that, as a virtue, courage is the proper balance between fear and confidence in the pursuit of the noble, which must occur at the right time, in the right place, and in the right way. They might also discuss the purpose of political

association and the means to achieve it, thereby specifying a hierarchy of ways of being a polity that can ground a hierarchy of ways of being a citizen. In other words, they will engage in the hard business of first-order normative theorizing.

In sum, since our interlocutors are employing a normatively used concept, they will find themselves in the familiar situation where the normative parameters containing hierarchies of possibilities that are triggered by the use of such concepts can be potentially misleading. For disagreements often concern the hierarchy itself, rather than the concept; those concepts may reliably indicate the presence of hierarchies, but not the reasons behind them. Accordingly, disagreements concerning them should be treated separately. And if a method can contribute to making explicit normative disagreements that are otherwise implicit, one can expect it to foster the development of first-order normative theories that are better suited to achieve their aim, that is, to ground the systematic production of normative judgments, thereby promoting philosophical progress.

**4. Preview.**

In Parts II and III, I illustrate the discussion developed so far by providing an analysis of the concept of unfreedom and by applying the proposed methodological procedure to some disagreements over it. In Part II, I offer an analysis of the concept of unfreedom that highlights how ascriptions of unfreedom are context-sensitive, especially by requiring a hierarchy of ways of being an agent that helps select which limitations are unfreedom-realizing. In Part III, I apply the methodological procedure just introduced to some disagreements between liberals and republicans in contemporary political theory concerning the concept of unfreedom.

# PART II - UNFREEDOM IN CONTEXT

## 0. Introduction.

In this Part, I explicate the concept of unfreedom in terms of limitations on practical activity that limit individual human potential in agency; specify three domains around which such limitations tend to cluster (viz., the formation of the will, the exercise of the will, and option sets); argue that ascriptions of unfreedom are context-sensitive; and argue that context often provides a normative parameter that helps fix a conception of individual human potential in agency, which in turn helps select the limitations underpinning particular ascriptions of unfreedom.

**1. The Concept Of Unfreedom.**

In this and the next Parts, I will employ a concept of unfreedom derived from Christine Swanton's *Freedom: A Coherence Theory* (1992). In this work, she develops a theory of freedom grounded in a coherentist methodology that attempts to achieve wide reflective equilibrium (i.e. to cohere) in a multifarious sample of ascriptions of freedom (in her terminology, freedom judgments) constituted of what she calls, borrowing from Aristotle, endoxa, that is, the freedom judgments of the wise and the many. One of the results of this work is an explication of the concept of freedom, that is, an analysis of the concept of freedom in terms of the theoretical role assigned to it by the broader coherence theory. My immediate task will be to build upon this result to provide an explication of the concept of unfreedom that is, as far as possible, non-normative[14].

I propose the following explication of the concept of unfreedom: it designates limitations on practical activity that limit individual human potential in agency. This is the pattern underlying ascriptions of unfreedom: it is realized by the set of items that satisfy the condition of limiting practical activity in a way that limits individual human potential in agency. If this set contains completely disparate items, as it seems to, then unfreedom is a multiply realizable property[15]. In any case, the pattern captured in the explication above is the most important piece of the *conceptual* puzzle. Therefore, before presenting some unfreedom-realizing items, I will elucidate the condition that they must satisfy.

Swanton characterizes the notion of practical activity in terms of 'a process of focusing desires onto the world through intentional action' (p. 52), intentional action being action for which instrumental reasons can be given, that is, action for which an explanation can be given in terms of its goal(s)[16]. Two immediate implications are worth noting. First, the

---

[14] Except for the closing section of this Part, most of what follows is a rather latitudinarian reconstruction of Swanton's work. As such, I am solely responsible for the claims and interpretations offered here. Two major departures from her work have already been mentioned: first, my subject is unfreedom rather than freedom, and my goal is merely to explicate the concept of unfreedom rather than to provide a theory of unfreedom; second, in keeping with the argument developed in Part I, I deliberately insulate the explication from normative assumptions in order to provide, as far as possible, a non-normative concept of unfreedom. The issue concerning the role of hierarchies of possibilities in ascriptions of unfreedom will be resumed in the closing section, where it will be accounted for contextually rather than semantically.

[15] Swanton expresses a similar idea about the nature of freedom by describing it as a formal property that can be instantiated by multifarious material properties (1992, p. 36).

[16] Strictly speaking, Swanton endorses an Anscombian view of intentional action, according to which it is action for which a reason can be given simpliciter (p. 33). But the broader context, and her later discussion on intentions (pp. 55-8) (which she does not take to be a necessary condition for intentional action), suggest a narrower interpretation of this view, according to which intentional action is goal-oriented action, that is, a sequence of movements oriented towards the attainment of a goal. Hence the proposed analysis in terms of action for which *instrumental* reasons can be given (whether or not the agent is able to articulate such reasons).

notion of intentional action as goal-oriented action is permissive enough to admit a large range of entities as being capable of practical activity (e.g. babies), thereby not imposing, at the outset, major restrictions (e.g. language or reasoning) to the set of objects amenable to ascriptions of unfreedom. Second, the notion of intentional action as goal-directed action is stringent enough to exclude accidental actions from its purview, and thereby from the purview of practical activity, which in turn solves, at the outset, apparent conundrums in ascriptions of unfreedom. For example, does a high fence on the edge of a cliff render one unfree in any way? It does not render one unfree to fall off that cliff, since falling, by being accidental, is not an object of intentional action; however, it does render one unfree to practice skydiving from that cliff, since skydiving is an object of intentional action.

She then spells out the process that characterizes practical activity in the following way:

> Agents' desires provide the rationale for practical activity; their deliberations rank their desires in order of importance, indicate their means of fulfillment, and determine when, how, and in what order they are to be fulfilled; their deliberations terminate in judgments about the thing to do; their intentions set their minds to performing the appropriate actions now or in the future; finally, they try to perform the relevant actions, and, barring breakdowns, those actions are performed (p. 52).

This is a description of what practical activity looks like when ideally completed; however, practical activity can be, and often is, non-ideally completed, e.g. by lacking some of these stages or by having these stages related in a way distinct from the one described. Thus, one can act upon a desire without an accompanying practical judgment about the thing to do, or contrary to such a judgment; one can form the intention to act in a non-desired way; one can deliberate about one's options and act without an intervening intention or practical judgment about the thing to do, and so on. The important point is that there are at least five roughly distinguishable stages in practical activity - desiring, deliberating, forming practical judgments, forming intentions, and attempting to act[17] -, each of which can be variously and distinctively limited[18].

---

[17] See Swanton 1992, pp. 52-8. I will not expand on her discussion about these stages, since the quotation reproduced above conveys the gist of her conceptions, except for her conception of desire, which she distinguishes from idle wishes in that the latter either are not 'a logically proper object of attempt', or would not be an object of attempt even if the agent 'had the requisite capacities, opportunities, and knowledge' and 'no conflicting desire' (p. 54) rationalizing an alternative action precluding such an attempt. For present purposes, the distinctions provided above should suffice.

[18] For convenience, in what follows I refer to desires, practical judgments, and intentions as practical attitudes.

But not all limitations on practical activity are unfreedom-realizing. The explication of the concept of unfreedom provided above confines its application to the subset of limitations on practical activity that also limit individual human potential in agency. Two immediate implications are worth noting. First, unfreedom-realizing limitations only impinge upon individual human agents. This restricts the class of agents whose practical activity can be limited in an unfreedom-realizing way to the class of human agents, as opposed to, say, non-human animals, or institutions. Second, unfreedom-realizing limitations only impinge upon the potential of individuals in agency. This restricts the telos relative to which practical activity is limited in an unfreedom-realizing way to the potential of individual human beings, as opposed to, say, the potential of the human species, or the values of equality or justice, or the purposes of God or society[19].

Swanton characterizes the notion of individual potential in agency in terms of three dimensions: capacity, propensity, and flourishing (pp. 39-40). Capacity is the power to 'bring about states of affairs' (p. 39). Conceptions of individual potential in agency emphasizing this dimension tend to underpin less demanding ascriptions of unfreedom, such as those focusing on limitations to the availability or eligibility of options, rather than on limitations to a person's will, or an option's quality. Propensity is a person's 'likely behavior given opportunity and lack of constraint' (p. 39). Conceptions emphasizing this dimension also tend to underpin less demanding ascriptions of unfreedom, such as those focusing on limitations to the options a person would propend to, rather than on limitations to propensities and options of a certain quality. Flourishing is 'constituted by the satisfaction and development of those needs and capacities which, under good conditions[20], human beings characteristically desire to satisfy and develop, and whose development and satisfaction they enjoy under those conditions' (p. 42)[21]. Conceptions emphasizing this dimension tend to underpin more

---

[19] See Swanton 1992, p. 38.

[20] Swanton characterizes the notion of good conditions as the absence of evil conditions (e.g. 'anxiety, loneliness, severe stress, [...] terror, fear, coercion' [p. 42]), as not being necessarily identical with normal conditions, and as being referred to the optimal development of the *individual* (pp. 42-3).

[21] The flourishing dimension plays two important theoretical roles in Swanton's account: it is intended to ground the value of freedom, and it is intended to accommodate virtue-oriented conceptions of freedom. I will not expand on these theoretical roles, but I want to make a few remarks on the conception of flourishing just quoted. First, it is restrictive enough to exclude some manifestations of individual potential: after all, humans have the potential to be both saints and serial killers, but neither possibility falls within the purview of the needs and capacities that humans *characteristically* desire to satisfy and develop with enjoyment. Second, it is broad enough to admit individual variability: after all, human potential varies with individuals' natures, even when conditions are virtually identical. However, I think that virtue-oriented views are not easily accommodated by this conception of flourishing for two reasons: first, the needs and capacities related to virtue are not *characteristically* the object of desire and enjoyment; second, the conditions for the optimal development of the virtuous individual might very well be bad conditions.

demanding ascriptions of unfreedom, such as those focusing on the quality of the capacities, propensities, and options that are being limited.

So far, I have explicated the concept of unfreedom in terms of limitations on practical activity that limit individual human potential in agency: this is the pattern underlying ascriptions of unfreedom. One might complain that the resulting explication is unsatisfactory, on the grounds that it suffers from underspecification, that is, because it is of too low a resolution to be informative. Admittedly, unfreedom as explicated is a low resolution concept; but this results from the kaleidoscopic nature of ascriptions of unfreedom, which imposes a theoretical trade-off between resolution and accuracy. I will return to this issue in the last section of this Part.

Fortunately, however, ascriptions of unfreedom tend to cluster around three domains, the specification of which can render the analysis above more informative. These domains concern limitations on the formation of practical attitudes (desires, practical judgments, and intentions), limitations on the conversion of practical attitudes into action, and limitations on the objects of practical agency[22]. In what follows, I will specify these domains by providing a characterization of the relevant limitations, and by showing why they can underpin ascriptions of unfreedom. Then, I will argue that contextual information is often necessary to help determine the limitation(s) underpinning a given ascription of unfreedom, especially information about which conception of individual potential in agency is at play, which in turn, and crucially, often depends on an ideal of agency, that is, on a hierarchy of ways of being an agent[23]. In the jargon of the philosophy of language, I will argue that the extension of ascriptions of unfreedom is determined by the intension provided by the concept explicated above plus a contextual parameter plus a normative parameter providing a hierarchy of ways of being an agent.

---

[22] For convenience, in what follows I refer to the process leading to intentional action as the will.
[23] Once a conception of individual potential in agency is settled, the question about whether or not a given limitation is unfreedom-realizing becomes entirely empirical. The discussion to follow just provides a rough chart of unfreedom's conceptual territory in terms of three prominent landmarks.

## 2. Limitations On The Formation Of The Will.

Limitations on the formation of practical attitudes can limit individual potential in agency, and thereby be unfreedom-realizing.

Consider the following cases:

(1) Joe is unfree to live a normal life, despite constant medical reassurance that he has no illness impeding him from doing so: he is hypochondriac.

(2) The citizenry was unfree to resist public authorities' attempt to instill panic, since access to accurate information was successfully obstructed.

(3) Jane was unfree not to be his victim: she was gullible, and he was an ingenious con artist.

(4) The attorney argued that the defendant was unfree to act otherwise; after all, not even saints could be expected to remain serene under those circumstances.

These cases seem to felicitously ascribe unfreedom. In what follows, I characterize two kinds of limitations that might explain why: irrationality and inauthenticity. These kinds of limitations ordinarily underpin ascriptions of heteronomy, because both convey the sense that an agent's will is not self-ruled. I discuss them separately because irrationality concerns how practical attitudes are formed, whereas inauthenticity concerns where practical attitudes come from.

### 2.1. Irrationality.

Practical activity can be limited with respect to the way practical attitudes are formed. A major limitation in this regard is irrationality. Humans are capable of practical reasoning, that is, of linking practical attitudes (desires, practical judgments, and intentions) and explicit models of the world (beliefs) in a coherent and evidence-sensitive manner. Limitations on this capacity are commonly believed to limit self-ruling[24].

In order to see the intuition behind ascriptions of heteronomy based on irrationality, consider the following cases: the case of the schizophrenic whose practical attitudes are linked to false beliefs about being persecuted; the case of the ideologue whose practical attitudes are

---

[24] Since the relevant limitations here concern practical reason, rather than theoretical reason, it would be more accurate to label this kind of limitation Practical Irrationality. I avoid this term for reasons of convenience and simplicity.

linked to oversimplified beliefs that are insensitive to evidence, and who systematically subordinates descriptive beliefs to normative judgments; the case of the sleep-deprived person who lacks the resources necessary to cohere practical attitudes and beliefs. We can call these agents heteronomous because their practical activity is limited by factors rendering the formation of their practical attitudes irrational: a mental disorder, an ideological belief system, and a temporary physical condition, respectively.

I am not interested in specifying all the ways in which the formation of practical attitudes can be rendered irrational, that is, in which practical attitudes fail to be linked to beliefs in a coherent and evidence-sensitive manner. However, like Swanton[25], I think that a conception of irrationality that is suited to underpin ascriptions of heteronomy should present a few other features. First, it should be relativized to human capacities: in this way, if someone is limited in his grasp of otherworldly Forms or universal reason, this would not conceptually entail that he is heteronomous, since grasping them is not within the scope of ordinary human capacities. Second, it should be relativized to developmental stages: in this way, if children are limited in, say, their ability to cohere desires and beliefs, this would not conceptually entail that they are heteronomous, since that ability might not be a proper requirement for their developmental stage. Third, it should accommodate some degree of individual variability: in this way, even if someone fails to meet thresholds established in terms of normal distribution or maximization, this would not conceptually entail that he is heteronomous, since these thresholds might not be proper requirements for him. Fourth, it should accommodate the possibility that there are normative facts, or that there are hierarchies of ways of being, that can be the subject of reason, thereby accommodating limitations on this kind of reason. In other words, it should accommodate limitations to rationality of ends: in this way, it can allow for more robust ascriptions of irrationality-based heteronomy, such as those based on the view that there is a given class of ends appropriate for flourishing human beings.

Albeit somewhat impressionistic, these remarks on irrationality-based heteronomy should suffice for seeing why it can underpin ascriptions of unfreedom. In certain contexts, practical attitudes undisciplined by reason result in activity that is at best erratic (for example, practical activity that does not update in the face of error), which in turn can hinder the development of other capacities or propensities that depend on rational practical agency (for example, by limiting one's ability to make feasible and long-term plans). Moreover, if there

---

[25] See Swanton 1992, pp. 120-2.

are normative facts that can be apprehended by reason, then conceptions of individual potential in agency that include rationality of ends would regard limitations to it as unfreedom-realizing. Therefore, to the extent that irrationality designates a limitation on an important human capacity (viz. of rational practical agency) and on the apprehension of hierarchies of ways of being, it can underpin ascriptions of unfreedom. Consequently, one can expect conceptions of individual potential in agency emphasizing the dimensions of capacities and flourishing to prevail in ascriptions of unfreedom based on irrationality.

## 2.2. Inauthenticity.

Alternatively, practical activity can be limited with respect to the origin of practical attitudes. A major limitation in this regard is inauthenticity, which can be of two types: vicariousness and dissociativeness[26]. Both types comprise a variety of ways in which an agent can consider his will alien to himself, or his practical attitudes as having an alien origin; as it were, he is not their author. Vicariousness refers to externally determined practical activity, whereas dissociativeness refers to inauthentically internally determined practical activity.

### 2.2.1. Vicariousness.

Practical activity that is externally determined, or vicarious, is considered inauthentic because the resulting practical attitudes have a foreign origin. A person acting under hypnosis is a standard example here: his actions will probably be reflective of another person's desires, deliberations, practical judgments, or intentions. Swanton also discusses less dramatic cases of vicariousness that are worth mentioning.

The first case is that of manipulation, or external deceptive influence on one's practical activity. This usually involves the attempt at shaping an addressee's attitudes and behavior by means of misrepresentation. When someone is unaware of being influenced in this way, or of its deceptive nature, or of its content, he is in a situation that is relevantly different from the one that would obtain were he aware of this influence. The relevant difference, of course, lies in the fact that successful manipulation externally limits the capacity of a person being the source of his own practical attitudes. To this extent, manipulation can be said to render practical activity vicariously inauthentic.

The second case is that of persuasion, or external influence on one's practical attitudes that bypasses deliberation. This usually involves the attempt at shaping an

---

[26] In Swanton's terminology, non-inner-directedness.

addressee's attitudes and behavior by non-rational means, especially by emotional means. In contrast with, say, advising, which is external influence designed to improve an addressee's formation of practical attitudes by means of deliberation, persuasion usually targets one's reactive responses in a way that restricts deliberation. Here, the relevant limitation is similar to those limitations underpinning ascriptions of heteronomy based on irrationality; however, the relevant limitation is not irrationality *per se*, but rather the fact that successful persuasion externally limits the capacity of a person being the source of his own practical attitudes[27]. To this extent, persuasion can be said to render practical activity vicariously inauthentic.

The third case is that of indoctrination, or external influence on one's practical attitudes that renders them insensitive to reality. This usually involves the attempt at shaping an addressee's attitudes and behavior through the inculcation of specific models of how the world is and how to act that are insensitive to reason, evidence, or both. In contrast with, say, teaching, the typical result is a limitation, rather than an expansion, on addressees' 'ability and motivation to form their wants and practical judgments on the basis of [adequate knowledge and reason]' (p. 129)[28]. Here, again, the relevant limitation is similar to those limitations underpinning ascriptions of heteronomy based on irrationality; however, the relevant limitation is not irrationality *per se*, but rather the fact that successful indoctrination externally limits the capacity of a person being the source of his own practical attitudes[29]. To this extent, indoctrination can be said to render practical activity vicariously inauthentic.

In sum, to the extent that hypnosis, manipulation, persuasion, and indoctrination externally limit, albeit in different ways, the capacity of a person being the source of his own practical attitudes, they can be said to render practical activity vicariously inauthentic. More colorfully, vicariousness encompasses those phenomena in which a person's practical activity is hijacked by a third party, or in which a person's will serves as a proxy for another person's will. Therefore, to the extent that vicariousness-based heteronomy designates a limitation on an important human capacity (viz. of being the source of one's own practical attitudes), it can underpin ascriptions of unfreedom[30]. Moreover, one can expect all conceptions of individual

---

[27] After all, not being the source of one's own practical attitudes is compatible with being persuaded of more rational practical attitudes.

[28] Unlike Swanton, who thinks that indoctrination is a formal property that resists being defined in terms of characteristic methods, contents, or aims (see pp. 126-9), I think indoctrination has a 'programming' aspect definable in terms of a characteristic method and content that should be considered conceptually necessary. However, this minor disagreement is inconsequential for present purposes.

[29] After all, not being the source of one's own practical attitudes is compatible with being indoctrinated with more rational practical attitudes.

[30] However, vicariousness-based heteronomy does not necessarily entail unfreedom: the inert nihilist who gets indoctrinated and thereby can at least have some (externally determined) practical judgment upon which to act,

potential in agency to rely on vicariousness in ascriptions of unfreedom, albeit probably in different degrees, and for different reasons (e.g. conceptions emphasizing the dimension of flourishing will probably relativize the unfreedom-realizing potential of vicariousness to the degree to which it prevents self-mastery and the pursuit of proper ends).

### 2.2.2. Dissociativeness.

Practical activity that is internally determined can be considered inauthentic, and therefore dissociative, when it involves a limitation in how a person relates to his own practical attitudes, especially his desires. The addicted gambler who has a desire not to gamble is a standard example here: he is dissociated from the obsessive desire that determines his practical attitudes towards gambling, perhaps because he does not identify with that desire. Since practical activity is the process of projecting desires onto the world, limitations rendering desires dissociatively inauthentic can be said to limit that process as a whole.

Which limitations render a desire dissociatively inauthentic? Swanton surveys some popular candidates in the philosophy of action, and shows that none necessarily entails unfreedom[31]. One popular candidate is impulsiveness, which denotes a limitation in a person's capacity to act rather than to merely react. The idea here is that people have the capacity of approving and endorsing as their own goals submitted to conscious scrutiny, and that influences undermining that capacity render practical attitudes inauthentic. Thus, the intemperate who is possessed by homicidal rage when he gets cut off in traffic, or the person suffering from OCD who cannot resist his obsessive thoughts or compulsive behaviors, can be said to have impulsively inauthentic desires.

Whether or not impulsiveness necessarily renders practical agency dissociatively inauthentic and therefore heteronomous, Swanton shows that it does not necessarily entail unfreedom, because renouncing the capacity to approve and endorse scrutinized goals may expand rather than limit individual potential in agency. She exemplifies this possibility with the person who dislikes buying and whose buying behavior follows a decision procedure that involves no deliberation, and with the temperate person whose mere emotive reactions are properly calibrated and based on the right desires (p. 131). In the first case, impulsiveness

---

or the cognitively impaired person who gets persuaded and thereby can at least have some (externally determined) coherent intention upon which to act, in some contexts can be said to be less limited in their capacities and propensities than they would be otherwise. Moreover, according to some conceptions of flourishing, strategies involving *pia fraus*, noble lies, and other manipulative influences are believed to expand rather than limit individual potential in some contexts.

[31] See Swanton 1992, pp. 129-35.

expands rather than limits a person's propensities, whereas in the second case impulsiveness expands rather than limits a person's flourishing, assuming a conception of individual potential as flourishing relativized to a normative ordering of desires.

Another popular candidate is non-identification, which denotes a limitation in a person's capacity to identify with the desires determining his practical agency. The idea here is that desires come in levels, that people have the capacity of identifying with meta-level desires as opposed to first-level desires, and that influences undermining that capacity render practical attitudes inauthentic. Thus, the addicted gambler who identifies with a meta-level desire not to gamble but whose practical attitudes are determined by a first-level desire to gamble, can be said to have inauthentic desires because he does not identify with the desires determining his practical attitudes.

The problem with this account is of course explaining why non-identification with first-level desires simpliciter makes it the case that a desire is inauthentic. After all, meta-level desires may also be determined in a way that could be said to render them inauthentic, such as when they result from brainwashing, group membership, childhood trauma, mental disorders, etc. Moreover, meta-level desires with which one identifies may be considered inauthentic whereas the first-level desires with which one does not identify may be considered authentic. Thus, to borrow Swanton's example, the anorexic who identifies with an obsessive meta-level desire for slimness rather than with a first-level desire to eat, and whose practical attitudes are determined by the former rather than the latter, could be said to have inauthentic desires (perhaps because they result from a disorder) despite identifying with them.

Since explanations of inauthentic meta-level desires in terms of non-identification with higher meta-level desires run the risk of appealing to a vicious infinite regress in which increasingly higher meta-level desires are invoked to explain the inauthenticity of lower meta-level desires, alternative candidates have been proposed. One worth mentioning is non-integration between meta-level and first-level desires: limitations on the capacity to integrate desires on different levels would render them inauthentic. However, non-integration does not seem to necessarily entail dissociatively inauthentic desires: one's meta-level and first-level desires might be considered authentic *because* they are not integrated, such as when one has conflicting desires when faced with moral dilemmas.

Whatever the correct account of the relationship between non-identification, non-integration, dissociativeness, and inauthenticity turns out to be, Swanton shows that it

will not necessarily entail unfreedom, because limitations on the capacities to identify with and integrate desires may expand rather than limit individual potential in agency[32]. She exemplifies this possibility with the anorexic whose practical attitudes cease to be determined by the obsessive desire for slimness with which she identifies, and with the oppressed housewife whose first-order desires cease to be integrated with her meta-level desire of remaining in her situation. In the first case, non-identification expands rather than limits her potential insofar as healthy agency is more conducive to flourishing than unhealthy agency, whereas in the second case non-integration expands rather than limits her potential insofar as it can spur a change in her situation that is more conducive to the development of her capacities, propensities, and flourishing.

However, to the extent that dissociativeness-based heteronomy designates a limitation on an agent's capacity and propensity to relate with his practical attitudes as if they were his own, i.e. in a way that he can be considered their author, it can underpin ascriptions of unfreedom. Moreover, since this kind of limitations often presupposes a hierarchical ordering of desires, one can expect conceptions of individual potential in agency emphasizing the dimension of flourishing to prevail in ascriptions of unfreedom based on dissociativeness.

## 2.3. Conclusion.

Circling back to the cases presented at the beginning of this section, one can say that (1) to (4) seem to felicitously ascribe unfreedom because: (1) involves irrationality; (2) involves vicariousness-based inauthenticity, and perhaps irrationality; (3) involves vicariousness-based inauthenticity; (4) involves dissociativeness-based inauthenticity. Therefore, at a first pass, limitations on the formation of practical activity, or on the capacity of the will to organize, can explain the cases presented above. However, this is only a first pass explanation; as I will argue later, contextual information is often required to determine which limitations underpin particular ascriptions of unfreedom.

---

[32] See Swanton 1992, pp. 134-5.

### 3. Limitations On The Exercise Of The Will.

Limitations on the translation of practical attitudes into action can limit individual potential in agency, and thereby be unfreedom-realizing.

Consider the following cases:

(5) Joe was more unfree than he thought: his depression not only prevented him from finding a new partner, it prevented him from having any desire at all.

(6) Jane discovered she suffers from akinesia, a condition that renders her unfree to work, as it hinders her performance of many everyday tasks.

(7) He who is not brave is unfree.

(8) Despite being tempted by the devil himself, Jesus refused to become unfree.

These cases seem to felicitously ascribe unfreedom. In what follows, I characterize two kinds of limitations that might explain why: those causing the will to collapse, and those causing will and action to be incongruent[33]. Both kinds of limitations involve a mismatch between practical attitudes and action, the former by preventing practical attitudes simpliciter from being translated into action, the latter by preventing only a subclass of practical attitudes from being translated into action.

### 3.1. Collapse Of The Will.

Limitations preventing practical attitudes from being translated into action can lead to a collapse of the will. Following Swanton[34], I will illustrate this phenomenon by (briefly) describing limitations to practical activity involved in anomie, irresolution, and vacillation. These limitations render individuals agentially sterile, so to speak: if they can commit to any course of action whatsoever, it is in a fickle and unstable manner. This is why I say that they can lead to the collapse of the will, that is, to a diminished capacity to command action.

---

[33] Swanton calls both kinds of limitations executive failure (see ch 10). Admittedly, the boundaries between some limitations in this domain (e.g. anomie) and limitations on the formation of the will are diffuse. However, like Swanton (p. 136), I maintain a conceptual distinction between both domains: whether or not the will is heteronomous, it can also be limited in whether and how it gets translated into action. In other words, irrationality and inauthenticity are compatible with executive success, whereas the limitations to be discussed here, by designating executive failure, are not.
[34] See Swanton, 1992, pp. 155-8.

Anomie, understood as a state of unordered desires, limits practical activity to the extent that it prevents the formation of practical judgments and intentions that would be necessary for the practical process to be translated into action. Desires provide practical activity with a telos relative to which judgments about the thing to do and intentions are formed. By being unable to order their desires, anomic individuals typically fail to complete their practical process in this way. Such a state can manifest itself in a variety of ways, as in apathetic inertia, anxious paralysis, impulsiveness, and also in irresolution and vacillation. Its distinctiveness, however, lies in it being a limitation occurring at the beginning of the practical process that can be variously manifested in other stages.

Irresolution, understood as a state in which practical judgments do not terminate in intentions, limits practical agency to the extent that it severs the link (i.e. the formation of intentions) between practical attitudes and action. By being unable to set out to do something, that is, by being unable to envisage a behavioral sequence that would allow for the translation of practical attitudes into action, irresolute individuals, despite having formed a judgment about the thing to do, typically fail to complete their practical process. For example, the writer who has a deadline to deliver a manuscript, has a desire to deliver it, forms the judgment that writing the manuscript is the thing to do, but fails to form an intention to sit down and write, either now or in an explicitly planned future, will probably procrastinate and fail to complete the practical process in a way that would allow him to deliver the manuscript by the deadline.

Vacillation, understood as a state of switching between practical judgments, limits practical agency to the extent that it prevents the formation of intentions that are stable enough to allow for the undertaking of successful action. Unlike irresolute individuals, vacillating individuals keep switching between judgments about the thing to do; like irresolute individuals, vacillating individuals typically fail to form intentions that can be translated into a determinate course of action. For example, the writer who switches between the judgments that writing the manuscript now is the thing to do and that answering his emails now is the thing to do, by being unable to abide by a judgment about the thing to do now, perhaps because he believes that 'reason should provide a clearly correct and hence stable ordering [of options]' (p. 157) when there is none, will probably fail to complete the practical process in a way that would allow him to translate either judgment into action.

To the extent that the limitations involved in anomie, irresolution, and vacillation result in inaction, they can be said to limit the realization of individual potential in agency,

whether it is conceived in terms of propensities, capacities, or flourishing. Therefore, they can, albeit often do not, underpin ascriptions of unfreedom.

### 3.2. Incongruence With The Will.

Limitations preventing certain practical attitudes from being translated into action can lead to an incongruence with the will. Here, the relevant limitations do not lead the will to collapse before it can be translated into action; rather, they can lead to a different kind of mismatch in which a given subclass of practical attitudes fails to be translated into action. These limitations render individuals agentially deviant, so to speak: if they commit to a course of action, it is in a manner that deviates from that subclass of practical attitudes. Of particular importance here are those limitations preventing the subclass of practical attitudes grounded in the agent's normative judgments from being translated into action[35]. In what follows, I briefly discuss Swanton's account of limitations of this type, namely, those limitations involved in weakness of will[36][37].

Swanton characterizes weakness of will as 'weakly acting against a perceived requirement' (p. 138). She elaborates on this characterization by rejecting the (then) prevailing assumption that weakness of will is a form of irrationality, in the sense that it contradicts an all-things-considered practical judgment. Instead, she thinks that weakness of will is best characterized as a phenomenon that arises because the options available to the agent are both rational and incompatible: given the same set of reasons, practical reason supports both the practical judgment that $P$ is the thing to do, and the practical judgment that $\neg P$ is the thing do. In such situations, the agent is faced with a dilemma, or aporia, to which weakness of will is a kind of response. Before specifying what constitutes this kind of response, it is useful to expand on how practical reason can give rise to practical judgments that are both rational and incompatible.

---

[35] Does this characterization entail a normative account of limitations of this type? No, because the relevant normative judgments here are those espoused by the agent, whereas the ascription of this kind of limitation does not necessarily involve any normative judgment. The situation here is similar to the one that obtains in belief reports, where belief reports do not necessarily entail any attitude (of belief or disbelief) towards the belief being reported. Similarly, ascriptions of, say, weakness of will report on an incongruence between the practical attitudes grounded in the agent's normative judgments and his actions, without necessarily entailing that the ascription itself conveys a normative judgment. Therefore, I disagree with the claim that a 'judgment that an agent was weak-willed on a particular occasion is a normative judgment' (p. 152), if it is understood as a conceptual claim.

[36] See Swanton 1992, pp. 137-52

[37] Weakness of will is sometimes discussed under the label of akrasia. More generally, akrasia can be understood as lack of control in practical activity, but this understanding is too underspecified for my purposes, since it does not differentiate between the domains of formation and exercise of the will. In order to avoid confusion, I do not discuss akrasia directly here.

Swanton distinguishes between two types of dilemmas, or aporias, of practical reason to which weakness of will can be a kind of response. The first involves goods whose utilities are not a linear function of the utilities of the individual actions that are necessary to ensure the attainment of those goods. The dilemma arises when the 'marginal contribution of *each* act to the production of a good may be sufficiently insignificant to be outweighed by the costs of that act' (p. 140, original emphasis), but the utility of the final good is greater than the individual utilities of the actions necessary to ensure its attainment. In such cases, practical reason supports both the practical judgment that taking action A is the thing to do, and the practical judgment that not taking action A is the thing to do.

In some cases, the relevant good is directly aimed at. For example, the good of cardiorespiratory fitness requires regular exercise, but the costs of going to the gym on a given occasion might exceed its benefits (say, because it is a snowy Sunday). On this occasion, the agent who desires the good of cardiorespiratory fitness can form the following practical judgments on the basis of the same set of reasons: (i) going to the gym is the thing to do, (ii) not going to the gym is the thing to do.

In other cases, the relevant good is indirectly aimed at. For example, an athlete who already enjoys the good of cardiorespiratory fitness, and whose set of reasons would appear to dictate the practical judgment that not going to the gym is the thing to do on a given occasion, might nonetheless face a dilemma because he aspires to what Swanton calls motivational integrity, which is a 'state of affairs in which desires that provide merely nonevaluative reasons[38] for action are extinguished or at least reduced in motivational force' (p. 144). Here, the agent distinguishes between normative and non-normative reasons in a given set of reasons, and practical judgments dictated by the former might conflict with practical judgments dictated by the latter. Thus, an athlete who aspires to the indirect good of motivational integrity can form the practical judgment that not going to the gym is the thing to do because he already enjoys the good of cardiorespiratory fitness and wants to avoid the discomfort of going out in a snowy Sunday, but he can also form the practical judgment that going to the gym is the thing to do because he does not want to be motivated by non-normative reasons, such as 'avoiding discomfort', wanting instead to be motivated by

---

[38] Swanton characterizes evaluative reasons for action as those pertaining 'to the evaluation of the object of a desire, whereas a nonevaluative reason for action pertains to the suffering of the agent caused by nonsatisfaction of a desire' (p. 142). This is not to be confused with the account of evaluative judgments I provided in Part I, since Swanton's characterization is more restrictive (e.g., I admit that what she calls nonevaluative reasons may ground normative judgments). I will follow her characterization in this section, but will talk of normative and non-normative reasons instead, which seem more apposite terms given her overall picture.

normative reasons, such as 'it is good to overcome discomfort'. The problem, as before, is that the indirect good of motivational integrity is not a linear function from the individual actions necessary to secure it; hence the dilemma.

The second type of dilemma, or aporia, of practical reason to which weakness of will can be a kind of response involves what Swanton calls strong evaluation, which 'is a "background" judgment about the kind of dispositions one wishes to have and to cultivate' (p. 151). Here, there is a narrower set of reasons organized around a judgment about the character one aspires to, or around an ideal of agency, that supports practical judgments that might conflict with practical judgments supported by the broader set of reasons: 'dilemmas that occur because the agent has *not yet* acquired the relevant desired character or dispositions' (p. 151, original emphasis). For example, a mother who is excessively compassionate, but wants to be the kind of person that promotes autonomy in her child, may, on a given occasion (say, in a low-risk situation where the child tries to climb a tree), rationally form the following practical judgments: (i) helping the child is the thing to do, (ii) not helping the child is the thing to do. The former judgment is supported by the broader set of reasons including the disutility resulting from her discomfort in not acceding to her compassionate inclinations, whereas the latter judgment is supported by the narrower set of reasons grounded in her ideal of agency[39].

Swanton argues that the indeterminacy generated by both types of dilemmas, or aporias, of practical reason can be solved by the formation of, and the identification of the will with, strong evaluations. Since utilitarian calculation does not yield incontrovertible all-things-considered practical judgments in such situations, perhaps because they involve non-reducible or incommensurable goods, the agent can appeal to a different strategy guided by an ideal of agency. 'In these kinds of situations, the rational agent acts in accordance with the following rule: act in line with a relevant strong evaluation (...) unless there is an "emergency"' (pp. 151-2). Appeal to such a strategy depends, of course, on the specific situation, since not all situations might be relevant or frequent enough to require it.

In view of the above, Swanton accounts for weakness of will as a kind of response to situations in which practical reason supports incompatible practical judgments, the agent

---

[39] Here, strong evaluations, or ideals of agency, are relativized to the agent. As a consequence, this indeterminacy of practical reason can afflict both 'low-minded' - for example, a psychopath, or someone who does not aspire to the good of motivational integrity, as defined above - and 'high-minded' agents - for example, someone who aspires to a noble character.

forms strong evaluations with which he identifies his will, and there is no genuine emergency: it is the kind of response in which the agent takes the 'soft option', or translates into action practical judgments that are expedient but contrary to practical judgments dictated by his strong evaluations, or his ideals of agency. Such actions are considered *weak* because they instantiate the 'soft option' when there is no genuine emergency, and are considered against one's *will* because they do not instantiate the option grounded in the agent's strong evaluations[40]. In brief, weakness of will is 'action against strong evaluation' (p. 149), absent genuine emergency.

Whether or not this account of weakness of will is sound[41], it nicely illustrates those limitations preventing certain practical attitudes - in this case, the subclass of practical attitudes grounded in the agent's normative judgments[42] - from being translated into action. To the extent that these limitations result in deviant action, they can be said to limit the realization of individual potential in agency, especially if it is conceived in terms of flourishing. Therefore, they can, and curiously often do, underpin ascriptions of unfreedom.

### 3.3. Conclusion.

Circling back to the cases presented at the beginning of this section, one can say that (5) to (8) seem to felicitously ascribe unfreedom because: (5) involves a collapse of the will, perhaps based on psychologically rooted anomie; (6) involves a collapse of the will, perhaps based on physiologically rooted irresolution[43]; (7) involves weakness of will; (8) involves weakness of will. Therefore, at a first pass, limitations on the exercise of practical activity, or on the capacity of the will to command, can explain the cases presented above. However, this

---

[40] Alternatively, if the agent instantiates the 'soft option' when there is no genuine emergency, but does not identify his will with strong evaluations dictating opposite practical judgments, the action could be considered merely weak.

[41] I think that this is a sound account of ordinary ascriptions of weakness of will, but I reject as unsound the argument for the claim that it is necessarily incompatible with rationalist accounts of weakness of will. Swanton successfully shows that it is incompatible with accounts based on traditional conceptions of rationality, which tend to emphasize instrumental reason, but is less unequivocal about alternative accounts based on more demanding conceptions of rationality that include rationality of ends, or normative reason.

[42] Another example of this class of limitations is what Swanton calls moral cowardice (pp. 157-8), which differs from weakness of will only in that the agent identifies with practical judgments incompatible with those dictated by *his* strong evaluations. For example, the tenured professor who faces ideological encroachments on his syllabus may form the following practical judgments: (i) opposing ideological encroachments is the thing to do, or (ii) not opposing ideological encroachments is the thing to do. If he identifies with (ii) and goes along with ideological encroachments on his syllabus - perhaps because his broader set of reasons includes fears of social disapproval, of being mobbed, ennui, etc., that counterweight his normative reasons (e.g. that he has a duty to promote education instead of ideological indoctrination) -, he might be displaying moral cowardice. Unlike weak-willed agents, this professor does not identify with (i) then acts on (ii) by appealing to non-genuine emergency; he identifies with (ii) because he overweights his non-normative reasons.

[43] I selected these examples to show that not only weakness of character can be at the root of unfreedom-realizing limitations pertaining to this domain, as sometimes Swanton seems to suggest.

is only a first pass explanation; as I will argue later, contextual information is often required to determine which limitations underpin particular ascriptions of unfreedom.

**4. Limitations On Option Sets.**

Limitations on the objects of practical activity can limit individual potential in agency, and thereby be unfreedom-realizing.

Consider the following cases:

(9) Joe was unfree to access the college campus without presenting proof of vaccination.

(10) The prisoner was unfree, despite his willingness to remain locked in his cell.

(11) Jane is so famous that she is unfree to go to the grocery store undisturbed.

(12) Joe and Jane had the same two options, but his choice was more unfree than hers because one of his options was accompanied with a threat of punishment.

These cases seem to felicitously ascribe unfreedom. In what follows, I characterize two kinds of limitations on option sets that might explain why: those rendering at least one option in a given set unavailable, and those rendering at least one option in a given set ineligible, all else being equal.

**4.1. Unavailability.**

One way in which a restraint can limit an option set is by rendering at least one option in this set unavailable, all else being equal. It is useful to begin by establishing some terminological groundwork. For present purposes, I define an option as any possible course of action, an available option as any possible course of action in a given world, and an unavailable option as any impossible course of action in a given world. Thus, since in human worlds flying like Superman is not a possible course of action, it is an unavailable option.

One can flesh out these notions in various ways[44]. Regarding human worlds (i.e. possibilities concerning humans), I prefer to flesh out the notion of an unavailable option in the following way: an option X is unavailable iff, were an agent willing to perform X, he would not be capable of performing X. The counterfactual condition of an agent willing to perform X is to be understood strongly, that is, as the condition of not having his will limited

---

[44] Swanton proposes that 'an action is unavailable to an agent [...] if and only if that agent himself cannot bring it about that he perform that action' (p. 92). Since there is controversy as to whether one can do what one cannot form a will to do, I prefer the counterfactual gloss offered in this paragraph.

in the ways discussed so far, that is, in its formation or exercise. The notion of being capable of performing X is to be understood as a function of abilities and opportunities.

Some consequences of this characterization are worth noting. First, the counterfactual condition renders the concept of unavailability more 'objective' by restricting its application to cases where 'internal' limitations to the formation or exercise of the will are absent. As a further consequence, unfreedom-realizing limitations involved in irrationality, inauthenticity, weakness of will, anomie, and so on, are excluded from ascriptions of unfreedom based on unavailability of options.

Second, the notion of capability as a function of abilities and opportunities renders the concept of unavailability more germane to ordinary ascriptions of unavailability, which tend to imply that inability is neither necessary nor sufficient for unavailability[45]. One can see that inability is not necessary for unavailability in cases where lack of opportunity makes it the case that an option is unavailable, despite there remaining an ability to perform X. For example, player A might be able to defeat Magnus Carlsen in the World Chess Championship, but this option becomes unavailable (to player A and everyone else) as Magnus Carlsen decides not to defend his world chess championship title. And one can see that inability is not sufficient for unavailability in cases where inability is indexed to time or persists at an individual level. When inability is indexed to time, one might not be able to perform X at t' without making it the case that X is unavailable. For example, you might not be able to play the piano at t', but were you willing to play the piano at t'' and given appropriate opportunities, you could develop the ability to play the piano at t''. That is, your inability to play the piano at t' does not make it the case that playing the piano is an unavailable option to you. Alternatively, when inability persists at an individual level, one might not be able to perform X unassisted without making it the case that X is unavailable. For example, you might not be able to become president in an electoral democracy by self-appointment, but were you willing to become president and given appropriate opportunities, you could attempt to obtain others' contribution to your endeavor. That is, your persistent inability to become president at an individual level does not make it the case that becoming president is an unavailable option to you.

I can now discuss why limitations rendering options unavailable can underpin ascriptions of unfreedom. Swanton argues that limitations rendering options unavailable can

---

[45] See Swanton 1992, pp. 92-5.

underpin ascriptions of unfreedom because they remove such options from the set of potential objects of practical deliberation, thereby limiting individual potential in practical agency. I think this explanation is partially correct, as long as the set of potential objects of deliberation is understood to be independent of an agent's capacity to deliberate, since the latter is excluded by the counterfactual condition of the definition of unavailability provided above. But a more complete, and simpler, explanation is that limitations rendering options unavailable can underpin ascriptions of unfreedom because an agent cannot propend, exercise a capacity, or flourish in the direction of such options.

This explains why one could not be unfree to fly like Superman, although flying like Superman is an unavailable option for humans. This general unavailability is not within the scope of unfreedom-realizing unavailabilities simply because it does not amount to a limitation on individual human potential in practical agency. Swanton contrasts this explanation with Berlin's explanation of similar cases: '[...] Berlin is correct to say that the inabilities "to fly like an eagle or swim like a whale" are not limitations on human freedom. He is right, though, not because unfreedom does not include inability, but because those types of ability are not within the range of *human* potentiality' (p. 98, original emphasis).

One consequence of this explanation is that the source and the nature of a limitation become irrelevant for ascriptions of unfreedom based on *unavailability of options*. Whether or not a restraint results from human agency, non-human agency, intentional agency, non-intentional social arrangements, natural events and so on, does not alter its limiting effect on the options available in an option set. Moreover, this explanation is compatible with there being restraints that simultaneously limit and expand an option set. For example, games and laws create possible courses of action by rendering impossible some courses of action: you must abide by the rules of chess to be able to play chess, and you must satisfy legal conditions to be able to enjoy political rights[46].

### 4.2. Ineligibility.

Another way in which a restraint can limit an option set is by rendering at least one option in this set ineligible. The intuition behind ascriptions of ineligibility is that a restraint R on an option X makes it harder to perform X. Regarding human worlds (i.e. possibilities concerning humans), I prefer to flesh out the notion of an ineligible option in the following

---

[46] Note, however, that my focus are limitations on a given option set that render at least one option unavailable, all else being equal - that is, limitations that have no further effects on the option set.

way: an option X is more ineligible iff, were an agent willing to perform X, the present or future costs of him performing X are higher with R than would be without R, all else being equal. Again, the counterfactual condition of an agent willing to perform X is to be understood strongly, that is, as the condition of not having his will limited in the ways discussed so far, that is, in its formation or exercise. The notion of present or future costs of performing X being higher with R than would be without R, all else being equal, is to be understood as the additional loss attached to the performance of X at the actual world as compared to the loss attached to the performance of X at a counterfactual world that is identical to the actual world, except for the absence of R.

Some consequences of this characterization are worth noting. First, ineligibility presupposes availability, that is, that the option be a possible course of action in a given world. More specifically, since the characterization is restricted to human worlds, for an option to be rendered ineligible the agent must be capable, in the sense explicated above, of performing it.

Second, the counterfactual condition renders the concept of ineligibility more 'objective' by restricting its application to cases where 'internal' limitations to the formation or exercise of the will are absent. As a further consequence, unfreedom-realizing limitations involved in irrationality, heteronomy, weakness of will, anomie, and so on, are excluded from ascriptions of unfreedom based on ineligibility of options[47].

Third, ineligibility is a gradable notion: an option X is more or less ineligible in proportion to the present or future costs of performing it. Note that the relevant class of comparison is the same option minus the restraint R imposing present or future costs[48], rather than distinct options.

Fourth, this characterization admittedly departs from ordinary ascriptions of ineligibility, which tend to encompass both 'external' and 'internal' limitations to agents'

---

[47] I will return to this consequence in the section where I discuss the kind of ineligibility that results from restraints imposing future costs.

[48] Cf. Swanton, 1992, p. 112. One implication of restricting the class of comparison in this way is to exclude opportunity costs, i.e. costs associated with foregoing an option, from the purview of ineligibility. Certainly, one way of influencing an agents' will is by manipulating the benefits attached to an option so that the costs of foregoing it become salient to him. Indeed, this can underpin ascriptions of unfreedom, such as when an addict in recovery is offered the substance he is addicted to. However, the unfreedom-realizing limitation here would concern the formation or conversion of his will, not his option set, which in fact would be unfortunately expanded.

actions, and to involve comparison between distinct options. However, I think the theoretical upside outweighs the ordinary parlance downside.

Swanton argues that restraints rendering options more ineligible can underpin ascriptions of unfreedom because they remove such options from the set of potential objects of intention, thereby limiting individual potential in practical agency. I think this explanation is partially correct, as long as the set of potential objects of intention is understood to be independent of an agent's actual intentions, since the latter are excluded by the counterfactual condition of the definition of ineligibility provided above. But a more complete, and simpler, explanation is that restraints rendering options ineligible can underpin ascriptions of unfreedom because an agent cannot propend, exercise a capacity, or flourish in the direction of such options without incurring in the loss required for overcoming those restraints.

I will now turn to the differences between ineligibility resulting from restraints imposing present costs and ineligibility resulting from restraints imposing future costs.

### 4.2.1. Restraints Imposing Present Costs.

Restraints imposing present costs to the performance of X typically impede agents from performing X. Following Swanton (pp. 113-15), I would identify the degree of ineligibility resulting from this kind of restraints with the least cost that an agent 'would incur while overcoming $R$ in a performance of [X]' (p. 113), whether overcoming R involves removing, bypassing, or enduring R. For example, suppose that a fence restraints your access to a park. You might access this park either by destroying the fence (removing R), going round the fence (bypassing R), or jumping over the fence (enduring R). In this case, the degree of ineligibility of accessing the park is identical to the least cost of overcoming the fence (presumably, the cost of either jumping over or going round the fence).

Typically, options restrained in this way have the effect of impeding agents from performing them. However, ascriptions of unfreedom based on this kind of ineligibility do not depend on there being typical impeding effects, since such effects operate at the level of the formation and/or exercise of the will. Instead, restraints rendering options ineligible by means of present costs can underpin ascriptions of unfreedom to the extent that an agent cannot propend, exercise a capacity, or flourish in the direction of such options without incurring in the present costs required for overcoming those restraints. Moreover, this kind of ineligibility does not necessarily entail unfreedom: for example, short fences impose present costs that are

negligible to physically apt agents, thereby failing to limit their potential in agency; likewise, expensive products impose present costs that are negligible to wealthy agents, thereby failing to limit their potential in agency.

### 4.2.2. Restraints Imposing Future Costs.

Restraints imposing future costs to the performance of X typically deter agents from performing X. Unlike Swanton (pp. 116-17), I would identify the degree of ineligibility resulting from this kind of restraints with the least cost that an agent would 'incur while overcoming $R$ in performing [X - whether overcoming R involves removing, bypassing, or enduring R -], multiplied by the objective probability of $R$'s occurring' (p.116)[49]. For example, suppose that you are an airline pilot and that a heavy storm is predicted to occur in your flight route, thereby restraining your option of taking that flight route. You might take that flight route either by circumventing the area in which the heavy storm is predicted to occur (bypassing R), or by risking getting caught in a heavy storm (enduring R). In this case, the degree of ineligibility of taking that flight route is identical to the least cost of overcoming a heavy storm multiplied by the objective probability of it occurring.

Typically, options restrained in this way have the effect of deterring agents from performing them. However, ascriptions of unfreedom based on this kind of ineligibility do not depend on there being typical deterrent effects, since such effects operate at the level of the formation and/or exercise of the will. Instead, restraints rendering options ineligible by means of future costs can underpin ascriptions of unfreedom to the extent that an agent cannot propend, exercise a capacity, or flourish in the direction of such options without incurring in the future costs required for overcoming those restraints. Moreover, this kind of ineligibility does not necessarily entail unfreedom: for example, predicted mild rains impose future costs that are negligible to skilled and well-equipped pilots, thereby failing to limit their potential in agency.

This analysis of ineligibility resulting from restraints imposing future costs also applies to cases of promised future costs, the only difference being that the objective probability of the latter usually also depends on the promisor's will. Since the most

---

[49] Swanton identifies the degree of ineligibility resulting from this kind of restraints with the least cost that an agent 'anticipates may be incurred by him while overcoming $R$ in a performance of $A$, multiplied by the subjective probability of that cost's occurring' (1992, p. 116). Since I conceptualize ineligibility as a property of an option, rather than as an agent's beliefs about an option, I prefer to make it depend on objective probabilities. In this way, my analysis helps avoid confusion with other unfreedom-realizing limitations.

conspicuous cases of promised future costs are cases of threats[50], I will discuss the Paradigm Case of the highway robber and two variations of it - which I will call the Bluff Case and the Deaf Case - in order to show why ineligibility resulting from restraints imposing promised future costs can underpin ascriptions of unfreedom, and two implications of the present analysis of ineligibility, viz. that ineligibility does not depend on how the actual world is to unfold in the future nor on agents' beliefs.

Consider first the paradigm case of the highway robber that approaches you at gunpoint and says: 'Your money or your life!'. In this situation, your option of going on the highway with your money is restrained by the robber's threat of imposing upon you the future cost of being shot. In other words, the robber's threat renders your option of going on the highway with your money ineligible, and to the exact degree of the least cost of overcoming his threat - say, by disarming the robber (removing R), dissuading him from executing his promise (bypassing R), or risking being shot (enduring R) - multiplied by the objective probability of him executing his promise. The latter depends on many factors, the most relevant of which of course being the fact that armed people issuing threats tend to be willing to execute them. Therefore, threats can underpin ascriptions of unfreedom for the same reasons that ordinary future costs can underpin ascriptions of unfreedom, namely, because an agent cannot propend, exercise a capacity, or flourish in the direction of the ineligible option without incurring in the probable future costs required for overcoming R.

Consider now a variation of the paradigm case in which the robber does not intend to impose the promised future cost, that is, in which his threat is a bluff. Does the fact that the robber is bluffing render you less unfree to go on the highway? It does not seem so. This case illustrates the broader intuition that the actual bearing of costs is not necessary for ascriptions of unfreedom. The present analysis of ineligibility resulting from restraints imposing future costs helps explain why: since the degree of ineligibility is determined by the least cost an agent would incur in overcoming R multiplied by the objective probability of R's occuring, future costs render options ineligible in the present, independently of how the actual world is to unfold in the future. Therefore, the ineligibility involved in the Bluff Case can underpin ascriptions of unfreedom to the extent that the agent cannot propend, exercise a capacity, or

---

[50] See Swanton 1992, pp. 104-09.

flourish in the direction of the ineligible option without incurring in the probable future costs required for overcoming R[51].

Consider now a variation of the paradigm case in which you are deaf and the robber approaches you from behind in a way that you cannot register his threat. Does the fact that you cannot register his threat render you less unfree to go on the highway? It does not seem so[52]. This case illustrates the broader intuition that awareness of a cost is not necessary for ascriptions of unfreedom. The present analysis of ineligibility resulting from restraints imposing future costs helps explain why: since the degree of ineligibility is determined by the least cost an agent would incur in overcoming R multiplied by the objective probability of R's occuring, future costs render options ineligible independently of agents' beliefs. Therefore, the ineligibility involved in the Deaf Case can underpin ascriptions of unfreedom to the extent that the agent cannot propend, exercise a capacity, or flourish in the direction of the ineligible option without incurring in the probable future costs required for overcoming R.

These cases show that the proposed analysis of ineligibility resulting from restraints imposing future costs encompasses cases of promised future costs, and how it helps explain ascriptions of unfreedom. However, ascriptions of unfreedom based on this kind of ineligibility can be confusing, especially because they can, and often do, overlap with ascriptions of unfreedom based on limitations on the formation and on the exercise of the will[53]. This results from the fact that the typical effects of ineligibility themselves can underpin ascriptions of unfreedom, thereby operating as a confounding factor in ascriptions of unfreedom. Indeed, it is often the case that the point in modifying an option's degree of ineligibility, or its believed degree of ineligibility, is to explore the typical effects that this kind of ineligibility has on the formation and exercise of a person's will in order to influence that person's will.

---

[51] This case of a falsely promised future cost differs from cases of falsely predicted future costs, because the probability of the former occurring partially depends on the promisor's will, whereas the probability of the latter occurring does not depend on the predictor. For example, suppose that a person is led to the false belief that going outside became more ineligible as a result of an airborne pathogen that imposes high and likely future costs. Although this false belief might produce the deterrent effects that would obtain were there an airborne pathogen, the degree of ineligibility of going outside remains unaltered because there is no R (airborne pathogen). On the other hand, in the Bluff Case there is a clear R (threat of being shot) the probability of which occurring depends, among other things, on the will of the robber. The Deaf Case will make it clearer that degrees of ineligibility do not depend on agents' beliefs.
[52] If anything, that fact only prevents you from overcoming the restraint imposed by his threat by means other than enduring the promised future cost, such as removing (e.g. killing the robber) or bypassing (e.g. dissuading the robber) his threat. However, the probability of him imposing the promised future cost remains the same.
[53] Ascriptions of unfreedom based on this kind of ineligibility can also be confusing because they can, and often do, overlap with ascriptions of insecurity and with ascriptions of domination. I will return to this point in Part III.

More specifically, by modifying an option's degree of ineligibility, or its believed degree of ineligibility, one typically influences the formation of a person's will by rendering less likely that the ineligible option becomes the object of his practical attitudes (desires, practical judgments, and intentions), and one typically influences the exercise of a person's will by rendering less likely that his willingness to pursue the ineligible option becomes translated into action. These deterring effects, which usually take place in the form of fear, shame, and disgust[54], can therefore underpin ascriptions of unfreedom, especially by making a person's will heteronomous or ineffective. However, whereas deterring effects vary with a person's risk-tolerance, cognitive abilities, knowledge, strength of will and so on, an option's degree of ineligibility remains identical to the least cost the agent would incur in overcoming R multiplied by the objective probability of R's occurring. Therefore, although ineligibility often is instrumental to the production of limitations on the formation and conversion of the will by means of its typical effects, ascriptions of unfreedom based on the latter should not be confused with ascriptions of unfreedom based on the former.

### 4.3. Conclusion.

Circling back to the cases presented at the beginning of this section, one can say that (9) to (12) seem to felicitously ascribe unfreedom because: (9) involves ineligibility resulting from restraints imposing present costs, and perhaps unavailability; (10) involves unavailability; (11) involves ineligibility resulting from restraints imposing future costs; (12) involves ineligibility resulting from restraints imposing promised future costs. Therefore, at a first pass, limitations on the objects of practical activity, or on option sets, can explain the cases presented above. However, this is only a first pass explanation; as I will argue now, contextual information is often required to determine which limitations underpin particular ascriptions of unfreedom.

---

[54] Fear is, among other things, a proxy for *subjective probabilities* of future damage; shame is, among other things, a proxy for *subjective probabilities* of future social disapproval; disgust is, among other things, a proxy for the *subjective probabilities* of future contamination. As such, these psychic costs do not bear on ineligibility as defined above.

## 5. Context-Sensitivity In Ascriptions Of Unfreedom.

So far, I have explicated the concept of unfreedom in terms of limitations on practical activity that limit individual potential in agency, and specified three domains around which such limitations tend to cluster. Particular ascriptions of unfreedom are the result of a function combining the limitations at play in a given context[55]. For example, a person might be unfree to perform X both because X is rendered ineligible and because his will is rendered heteronomous as a result of indoctrination; or, a person might be rendered less unfree to perform X *because* X is rendered ineligible to a degree that X becomes enticing and helps discipline his will. I will now argue that the determination of these functions often requires contextual information, that is, I will argue that particular ascriptions of unfreedom are context-sensitive. Before doing so, however, I need to address a complaint raised at the beginning of this discussion.

I mentioned earlier that one might complain that the proposed explication of the concept of unfreedom is unsatisfactory because it is of too low a resolution to be informative. At this point, I hope that the preceding discussion has solved the issue of informativeness. However, the issue of low resolution remains: is it not theoretically advantageous to increase the resolution of the concept by restricting it to some subset of limitations (say, on option sets) or by decomposing it in dimensions corresponding to the domains specified above? In order to answer this question, the theorist must be aware that there is a trade-off between resolution and accuracy in analyses of unfreedom: increased resolution comes at the expense of decreased accuracy. Accordingly, restricting the concept of unfreedom to some subset of limitations would increase resolution by giving up elucidation of felicitous ascriptions of unfreedom based on other limitations, thereby decreasing accuracy; decomposing the concept of unfreedom in dimensions would increase resolution by begging the question of what makes them dimensions of the same concept, thereby potentially decreasing accuracy.

---

[55] This view lends some support to Côté's (2020) probabilistic approach to measuring freedom, which models option sets (in his terminology, opportunity sets) as fuzzy sets that admit many parameters in order to probabilistically represent how accessible a given option set is. The motivation behind this approach is to account for the intuition that freedom can vary in degree for otherwise identical option sets: for example, it allows to account for the intuition that, for identical option sets, agents endowed with more willpower are freer than less endowed agents by modeling their option sets as fuzzy sets whose accessibility is represented as a probabilistic function of willpower. I sympathize with the approach, but I think it has the tendency to turn ascriptions of freedom into black boxes that can be misleading by (i) obscuring more than elucidating why some parameters are believed to impinge on freedom; (ii) obscuring that distinct parameters can increase and decrease freedom at the same time; (iii) obscuring that the same parameter can decrease and increase freedom at different times - say, as when an agent's option set requires more willpower to be accessed, thereby rendering him less free in the present, but whose effortful access develops more willpower, thereby rendering him freer in the future (cf. Côté, 2020, pp. 403-4).

Surely, accurate correspondence with ordinary language is not a good criterion for conceptual analyses that intend to elucidate rather than strictly represent ordinary language. However, I think the current explication of unfreedom is preferable not because it covers more ordinary language cases, but because it uncovers an underlying pattern that provides a unified explanation of the kaleidoscopic nature of ascriptions of unfreedom. In other words, I prefer the current explication on Occamist grounds, and I endorse it by inference to the best explanation; hence, I land on the accuracy side of the trade-off just mentioned[56]. This is why I take the domains specified above not as dimensions, but as landmarks in an otherwise labyrinthine conceptual territory. In any case, even if the approach adopted here turns out to be excessively permissive, it is at least cautious enough to allow for a fair comparison between rival accounts of unfreedom, which will be one of the tasks of Part III.

I turn now to the claim that context helps select the limitations underpinning particular ascriptions of unfreedom. In order to see how, consider the following sentence in three different contexts:

(13) Jason was unfree to leave the building.

Consider this sentence in the following context: Jason is an addicted gambler who unfortunately finds himself in a casino. Presumably, this context would elicit the interpretation that something is amiss with Jason's will, that he is a slave of his passions. More specifically, context would make salient those limitations to Jason's option of leaving the building related to irrationality (e.g. his addiction being a limitation on his rationality of ends), dissociation (e.g. his addiction being a limitation on the authorship of his practical attitudes), or, less importantly, weakness of will (e.g. his addiction being a limitation on the alignment between his actions and his will). In other words, context helps select limitations on the formation (irrationality- or dissociation-based heteronomy) and, less importantly, on the conversion of Jason's will.

Now, consider this sentence in the following context: Jason is at home, but suffers from agoraphobia. Presumably, this context would elicit the interpretation that something is amiss with Jason's ability to act. More specifically, context would make salient those

---

[56] See Swanton 1992, ch. 2, for her methodological reasons behind her concept of overall freedom. It should be noted that, unlike her (see ch. 12), I refrain from attempting to cohere conflicting ascriptions of unfreedom, since my aim here is merely to explicate the concept of unfreedom rather than to provide a theory of unfreedom, and since my main aim in the next Part is precisely to illustrate the application of a method for properly diagnosing disagreements over ascriptions of unfreedom.

limitations to Jason's option of leaving the building related to dissociation (e.g. his agoraphobia being a limitation on the authorship of his practical attitudes), or, more importantly, to the translation of his will into action (e.g. his agoraphobia being a limitation causing his will to collapse in irresolution or vacillation). In other words, context helps select limitations on the formation and, more importantly, on the conversion of Jason's will.

Now, consider this sentence in the following context: Jason is held in custody in a police station for interrogation. Presumably, this context would elicit the interpretation that something is amiss with Jason's options, regardless of his will. More specifically, context would make salient those limitations to Jason's option of leaving the building related to unavailability and ineligibility. In other words, context helps select limitations on Jason's option set.

In sum, context helps select the limitations underpinning particular ascriptions of unfreedom by making some limitations more salient than others. Thus, and beyond the examples above, one can expect limitations on the formation of the will to be particularly salient in educational contexts; one can expect limitations on the conversion of the will to be particularly salient in psychological contexts; one can expect limitations on option sets to be particularly salient in political contexts; and so on. Put differently, context provides functions of ascriptions of unfreedom with weighted priors, thereby increasing the likelihood that those limitations assigned greater priors get selected to underpin ascriptions of unfreedom.

However, this is not the full story. Frequently, context also provides a normative parameter containing a hierarchy of ways of being an agent, which in turn helps fix a conception of individual potential in agency, which in turn helps select the limitations underpinning particular ascriptions of unfreedom[57]. For example, if context provides a normative parameter containing a hierarchy of ways of being an agent defined in hedonistic

---

[57] Swanton is not oblivious of the role normative systems play in ascriptions of unfreedom, which she registers by pointing out that ascriptions of unfreedom are also a function of the significance of options, i.e. that limitations on significant options render an agent more unfree than limitations on less significant options (1992, ch. 11). Moreover, she notes that distinct conceptions of individual potential in agency yield distinct conceptions of significance - i.e. for conceptions of potential as propensity, significance 'should be interpreted simply in terms of agents' wants' (p. 162), and for conceptions of potential as flourishing, significance 'should be connected with self-realization' (idem) -, and tries to provide an analysis of significance of options in terms of *interests* that (i) is not identical to agents' wants, (ii) is applicable to imperfect worlds, and (iii) accommodates the notions that agents are the ultimate arbiters of their interests and that agents' interests promote their good. She arrives at the following analysis: option X is in a person's interest if it is preferred to the alternatives were he to experience all options under conditions of integration, nondistortion, and detachment, and after 'all costs of acquiring, pursuing, and satisfying those wants are taken into account' (p. 176). Despite laudable, her attempt to strike a middle ground between conceptions of significance as interests departs from the thrust of the present discussion, which is precisely to avoid turning substantive normative issues into analytical issues.

terms, one can expect a conception of individual potential in agency as propensity to prevail, and, as a consequence, one can expect less demanding ascriptions of unfreedom to prevail - say, ascriptions focusing on limitations on option sets, or on the conversion of the will, regardless of the quality of such options and practical attitudes. Thus, if the contexts described above are supplied with a normative parameter informed by a hedonistic normative system, i.e. by a hierarchy of ways of being an agent defined in terms of pleasure and displeasure, the last scenario is more likely to be picked out as unfreedom-realizing, because Jason's addiction or agoraphobia are less likely to be seen as limitations to his propensities.

Alternatively, if context provides a normative parameter containing a hierarchy of ways of being an agent defined in virtue-oriented terms, one can expect a conception of individual potential in agency as flourishing to prevail, and, as a consequence, one can expect more demanding ascriptions of unfreedom to prevail - say, ascriptions focusing on limitations on the formation of the will, or on option sets, precisely on account of the quality of such practical attitudes and options[58]. Thus, if the contexts described above are supplied with a normative parameter informed by a virtue-oriented normative system, i.e. by a hierarchy of ways of being an agent defined in terms of self-mastery and proper ends, the first two scenarios are more likely to be picked out as unfreedom-realizing, because Jason's addiction or agoraphobia are more likely to be seen as limitations to his flourishing.

If this story about how context helps to fix the content of ascriptions of unfreedom is correct, then one can expect disagreements over ascriptions of unfreedom to be particularly susceptible to the problem explored in Part I, namely, to the problem of conflating verbal and normative disagreements. In the next Part, I will investigate whether this expectation holds for disagreements occurring between liberals and republicans in contemporary political theory.

---

[58] More precisely, I think the normative systems informing normative parameters in modern ascriptions of unfreedom are of a defectionist rather than perfectionist bent. That is, I think that when normative parameters help fix ascriptions of unfreedom, they are often informed by hierarchies of bad ways of being an agent (or by ideals of bad agency), rather than by hierarchies of good ways of being an agent (or by ideals of good agency). (Even though it could be reasonably argued that defectionism necessarily presupposes perfectionism, I think it is useful to distinguish between the two when describing hierarchies). I will set aside this interesting issue in the history of normative systems because it is inconsequential for the present discussion.

# PART III - UNFREEDOM IN THE POLITICAL CONTEXT

## 0. Introduction.

In this part, I apply the methodological procedure introduced in Part I to some verbal disagreements between liberals and republicans in contemporary political theory concerning the concept of unfreedom; argue that the disagreement over whether or not unfreedom requires interference should be diagnosed as merely verbal; argue that the disagreement over whether or not unfreedom only requires interference should be diagnosed as legislatively verbal; and present some normative disagreements that might underlie verbal disagreements over unfreedom.

## 1. Verbal Disagreements Over Unfreedom.

Political contexts usually make salient limitations on option sets and limitations on the formation of the will in ascriptions of unfreedom. This is nicely illustrated by the debate between liberals and republicans in contemporary political theory[59]. Central to this debate are disagreements over the concept of freedom, which liberals analyze in terms of non-interference, and republicans analyze in terms of non-domination. Given the negative construal of their analyses, one could properly reformulate those disagreements as being disagreements over the concept of unfreedom, which liberals would analyze in terms of interference, and republicans would analyze in terms of domination. In what follows, I will apply the methodological procedure introduced in Part I to some of these disagreements. I begin with a brief review of distinctions that are important for that methodological procedure.

In Part I, I characterized verbal disagreements as disagreements over a sentence S that result from disagreement over the usage of at least one term T in S. In other words, I said that parties verbally disagree over S when they are disagreeing over the meaning of at least one term employed in S. I then distinguished between two kinds of verbal disagreements, viz. merely verbal disagreements, which seem pointless, and legislatively verbal disagreements, which seem to have a point. Merely verbal disagreements were characterized as being devoid of representational significance, since they merely reflect contrasting word usage, and were predicted to disappear should the conflicting parties agree on the usage of T. Legislatively verbal disagreements were characterized as being of ulterior significance, since they reflect ulterior substantive disputes (e.g. linguistic, practical, legal, normative), and were predicted to persist in a distinctive way: the conflicting parties would not be able to agree on the usage of T precisely because their ulterior substantive disputes legislate contrasting roles to T. In order to sort out merely from legislatively verbal disagreements, I proposed either barring or replacing T: if the disagreement disappears, there is evidence for diagnosing it as merely verbal; if the disagreement persists, there is evidence for diagnosing it as legislatively verbal. Finally, I also said that, should the latter obtain, one could expect the ulterior substantive disputes legislating contrasting roles to T to be made more explicit, thereby preventing one form of dispute from being mistaken for another, which is itself a form of progress.

---

[59] In this Part, I will focus on the debate that unfolded in the past two decades between the main proponents of the so-called pure conception of negative liberty and the so-called neorepublican conception of liberty.

I will now present a major verbal disagreement between liberals and republicans. Liberals claim, and republicans deny, that unfreedom requires interference. Republicans often resist this claim with the following *reductio*:

Premise 1: Unfreedom requires interference.

Premise 2: Some slaves are not interfered with.

Conclusion: Therefore, some slaves are not unfree.

Republicans take this conclusion to be absurd, because it is incompatible with the fundamental folk intuition that all slaves are unfree. Since they take the truth of Premise 2 to be demonstrated by, say, cases of slaves that are not interfered with because they have benevolent masters, republicans reject the absurd conclusion by denying Premise 1, that is, by denying that unfreedom requires interference.

A slightly different, but more complete version of this argument goes as follows:

Pr. 1: Unfreedom requires interference.

Pr. 2: All slaves are unfree.

Pr. 3: There can be slavery without interference.

C. 1: Therefore, there can be unfreedom without interference.

C. 2: Therefore, unfreedom does not require interference.

Call it the Argument from the Lucky Slave[60] (ALS). Pr. 1 registers the liberal claim, Pr. 2 registers the fundamental folk intuition mentioned above, Pr. 3 registers the contingent truth demonstrated by benevolent master cases, C. 1 registers what the conjunction of Pr. 2 and Pr. 3 allegedly entails, and C. 2 registers that C.1 entails the denial of Pr. 1. Republicans take this argument to show that interference is not a necessary condition for unfreedom. Besides this negative claim, republicans also take this argument to support their positive claim that domination, or power to interfere arbitrarily, is a sufficient condition for unfreedom. This is how they explain why slaves in benevolent master cases remain unfree: whether or not a slave is interfered with, he is unfree because he is dominated. In sum, and in terms of

---

[60] The Lucky Slave label is borrowed from Goldwater 2020.

properties, republicans take this argument to show that the main unfreedom-realizing property in the political context is domination, rather than interference.

Before expanding on how the disagreement over whether or not unfreedom requires interference proceeds from the ALS, it is important to notice that specific disagreements between liberals and republicans occur against a background of some general agreements concerning the scope of unfreedom-realizing influences. First, both parties restrict this scope to human influences, thereby excluding, say, influences resulting from natural events. Second, both parties restrict this scope to influences that have limiting effects, especially on option sets, thereby excluding, say, offers that expand option sets. Third, both parties emphasize influences on compossible options, that is, on options that agents can enjoy in a compatible way. Fourth, both parties emphasize intentional influences. Fifth, both parties tend to admit that the greater the significance of an option, the greater the unfreedom-realizing potential of influences limiting it. Finally, both parties tend to admit that influences on agents' attitudes towards options, such as misrepresentation and intimidation, can be unfreedom-realizing. In what follows, I assume this general background agreement.

How does the disagreement over whether or not unfreedom requires interference proceed from the ALS? Roughly, it proceeds with liberals arguing, and republicans rejecting, that a proper understanding of their concept of interference can account for the putative counter-examples of unfreedom without interference.

Liberals could begin by saying that they are

concerned with the absence both of actual interferences and of a large class of possible-but-not-actual interferences. [...] First, liberals characterize unfreedom in terms of interferences that exist both actually and counterfactually, in more or less nearby possible worlds, conditionally upon a series of actions or omissions on the part of the agent whose freedom is under consideration […]. Second, for each action or set of actions, we must ask how probably constrained it would be in the event of the agent attempting to perform it. The liberal conception of freedom therefore takes into account all possibilities of interference starting from the actual world at the time of the freedom being assessed [...] but weighs each possible interference in terms of the probability of its occurrence (on the assumption that the agent attempts to perform the action that might be subject to that interference). The greater the probability of interference (conditional upon the agent attempting to perform the relevant action), the

more that interference contributes to the agent's unfreedom[61] (Carter & Shnayderman, 2019, pp. 137-8).

Equipped with this understanding of interference, liberals could then argue that 'the relation between the unfreedom of B and certain forms of power of A is [...] an empirical generalization' (Carter, 2008, p. 68), and account for the unfreedom involved in cases where a person is dominated by, or depends on, another in the following way:

> Though her state of dependence might eliminate very few of her particular freedoms, it will eliminate the conjunctive exercisability of many such freedoms and will thereby significantly reduce her overall liberty. Specifically, the combinations of conjunctively exercisable liberties that will typically be removed are those which include freedoms to engage in patterns of behavior that are not deferential toward the dominant party (Kramer, 2008, p. 35).

> Far from being a situation wherein interference is unlikely, P's plight is a situation wherein interference is occurring extensively. That interference does not come about through the actual application of violence (*ex hypothesi*), but it consists in the undoing of the conjunctive exercisability of many opportunities – opportunities that could have been exercised conjunctively in the absence of the dominant party's sway. In the presence of that sway, if P acts in any manner that is insufficiently humble or furtive, she will not also be able to act in any manner precluded by the retaliation that will be undertaken against her as a response to her perceived audacity. Her freedom to act in the former manner is not exercisable conjunctively with her freedom to act in the latter manner, because of the dominant party's preparedness to remove the latter freedom if the former freedom is acted upon. Hence, under any tenable account of negative liberty, we must conclude that P's overall freedom is substantially reduced by her subjection to the dominant party. [...] [T]he dominance of that party is indeed being exerted – not through the actual infliction of violence (which is unnecessary in the circumstances), but through the party's readiness to inflict violence. That very readiness eliminates many combinations of conjunctively exercisable freedoms for P' (Kramer, 2008, p. 44).

---

[61] Not all liberals think that 'freedom is impaired in proportion to the probability of someone interfering with your actions and choices' (Goodman & Jackson 2007, p. 250). See, for example, Lang 2012, p. 283.

Moreover,

> on the liberal conception of freedom, a slave of an interfering master is less free than a slave of a noninterfering master. This is not to say, however, that a slave of a non-interfering master is as free as someone who is not a slave. Even in the presence of a non-interfering master, a slave is normally much more likely than a non-slave to suffer future interferences (Carter & Shnayderman, 2019, p. 142).

In turn, republicans could reply in the following way:

> Suppose that B wants to raise the probability of A's choosing x. Absent perturbing factors, the best thing for B to do when A is disposed to choose y is to interfere, and the best thing for B to do when A is disposed to choose x is to stand by, guarding against a possible change of mind. That is exactly what B does. B controls A by active interference in the one case and by invigilation, i.e., invigilation with a view to possible interference, in the other.

> The control that B exercises over A's choice reduces A's freedom, intuitively, because it is an uncongenial form of control that affects A's can-do assumptions in respect of the original options, x and y. It makes one or another assumption false, whether by manipulating A's rationality, by removing one of A's options, or by replacing an option with a burdened alternative. Or at least it leads A to believe that one or another assumption is false. (Pettit 2008b, pp. 213-4).

Therefore,

> B's interference may become less probable without B's uncongenial control being in any way lessened. And in those cases it seems quite wrong to say that A's freedom has increased. Thus it is intuitively wrong to say that A enjoys a greater degree of freedom just because A's taste as between x and y shifts in B's direction or because B's taste for interfering with A happens to shift in A's favor. [...] A's freedom is a function of the probability of interference under the possibility associated with each option and is not sensitive to the probability that one or another option will be chosen. (Pettit 2008b, pp. 216-7).

Put comparatively,

The republican principle expands still further the range of possibilities that become nonprobabilistically relevant in the measure of freedom. [...] The [liberal] says that even if A suffers no interference in the actual world, A's freedom will still be sensitive to the probability of interference in the counterfactual world in which A chooses otherwise. The republican goes further still and says that even if A suffers no interference in either of those worlds, A's freedom will be sensitive to the probability of interference in the counterfactual worlds where B ceases to be charmed or duped by A (Pettit, 2008b, pp. 217-8).

In other words, B's power to interfere with A is unfreedom-realizing, regardless of the probability of it being exercised.

Liberals could retort that the probability, rather than the mere possibility, of B's interfering with A is enough to explain republicans worries:

where [B's][62] mere opportunity to exercise power has some degree of probability of being exercised, then [A's] unfreedom is to that same degree limited. And it would [...] be a very unrealistic theory of politics that conceived of opportunities for the exercise of power as being accompanied, except in rare cases, by a trivially low probability of that exercise taking place. Where [B's] capacity and opportunity to make threats is superior to [A's], but [B] does not exercise this superior coercive power, [B] usually exercises power over [A] thanks to the law of anticipated reactions. And in those cases in which [B] does not exercise power even in this way, the probability of those exposed to coercive power nevertheless coming to experience it is generally kept above a certain minimum by two factors: first, [B's] own behavior might change at any moment, for whatever reason; secondly (and perhaps more importantly), since power (as well as health and life itself ) is never fully guaranteed, [B] might at any moment be replaced as the holder of the relevant institutional position of power by some as yet unknown person. Thus, even if the power is not exercised by [B], the same power may still be exercised by [B'] (or [B''], or [B'''], . . . ) (Carter, 2008, p. 70).

Only with respect to the mind-bogglingly unrealistic scenario of a slave whose non-interfering master is (a) extremely unlikely to ever interfere with the slave's actions in the future in any of the hypothetical scenarios accessible to the slave, (b)

---

[62] The original labels 'A' and 'B' were switched to avoid confusion with other excerpts.

extremely unlikely ever to sell or give the slave to another person, and (c) extremely unlikely to die without having first provided for the manumission of the slave in the event that the master dies before the slave does—and despite all this, for reasons known only to herself, the master chooses not to free the slave within her own lifetime—does the liberal conception entail that, all other things being equal, a slave can be virtually as free as a non-slave. (Carter & Shnayderman, 2019, p. 142)

In short, in the very rare circumstances where relationships of domination genuinely involve extremely low probabilities of nontrivial encroachments on the freedom of subordinate people, we should not characterize the state of subordination as a state of unfreedom (Kramer, 2008, p. 49).

Finally, unmoved, republicans could restate their position in a somewhat different language:

Slaves are never free, because they are never free of their master's will; their actions are invariably performed by the leave and with the grace of someone else. As a result, a slave's pattern of conduct is nothing other than a reflection of what their master is willing to tolerate. This in turn means that, even if there is almost no probability that such slaves will be subjected to interference in the exercise of their powers, their fundamental condition of servitude remains wholly unaffected. It is the mere fact that their master or ruler has arbitrary powers to intervene that takes away their liberty, not any particular degree of probability that these powers will ever be exercised (Skinner, 2008, pp. 96-7).

A decrease in the probability of interference at the hands of an alien controller will not remove the specter of alien control (Pettit, 2008a, p. 124).

According to the republican tradition, the lack of liberty suffered by slaves is only secondarily due to there being various courses of action they may feel constrained to follow or avoid. [...] The basic condition in which slaves find themselves is not that some of their particular actions may be under some degree of pressure or threat; it is that they are condemned to living wholly at the mercy of their master's arbitrary power. [...] [T]hey are under alienating control, and it is this condition that leaves them bereft of liberty (Skinner, 2008, p. 99).

In sum, liberals often resist the ALS by saying that it involves a non sequitur, that is, that C. 1 (and consequently C. 2) does not follow from Pr. 1-3, properly understood. More

specifically, liberals accept Pr. 1-3, as long as (i) Pr. 1 is properly understood as including interferences in the actual world, interferences in possible worlds where the interfereé chooses differently, and probable interferences; (ii) Pr. 2 is properly understood as not being analytically true, but as being a generalization of the contingent fact that slaves face a higher probability of pervasive interference than non-slaves; and (iii) Pr. 3 is properly understood as being a possibility that should be assigned a low probability, since slaves face a higher probability of pervasive interference than non-slaves. Thusly understood, Pr. 1-3 would not entail C. 1 (and consequently C. 2), mainly because Pr. 2 is denied an analytical reading. In fact, since liberals explain the unfreedom involved in slavery in terms of interference (properly understood), they seem ready to drop the folk intuition that all slaves are unfree: if a slave is not interfered with in the actual world, would not be interfered with in those worlds in which he chooses differently, and does not face relevant probabilities of being interfered with, he would be a slave only nominally.

In turn, republicans reject that liberals can resist the ALS in this way. They insist that this line of reasoning fails to account for the persistence of unfreedom in cases satisfying liberals' proper understanding of interference. More specifically, republicans point out that interference in the actual world and in possible worlds where the interfereé chooses differently can be avoided, and its probability kept low, by expedients that attest, rather than obviate, unfreedom, e.g. by behaving in a servile and furtive manner. Republicans argue that, in order to account for the persistence of unfreedom in these cases, one must include in the scope of unfreedom-relevant possible worlds those in which the interferer chooses differently, independently of the interfereé's actual or possible choices. They propose to cover this widened scope with the concept of domination, or power to interfere arbitrarily, and they explain its unfreedom-realizing sufficiency in terms of its effects on the interfereé's will. More precisely, they claim that domination is a sufficient condition for unfreedom because it renders the interfereé's will a function of the interferer's will: thus, whether or not a slave is interfered with, his master's power to interfere arbitrarily renders him unfree because, ultimately, it is the master's will that determines the efficacy of the slave's will. Moreover, this explanation lends greater, not less, support to the folk intuition that all slaves are unfree. Republicans therefore conclude that interference is not a necessary condition for unfreedom, and that domination is a sufficient condition for unfreedom.

In my taxonomy, one could say that liberals diagnose the disagreement over whether or not unfreedom requires interference as merely verbal: since republicans' putative

counter-examples of unfreedom involve interference, properly understood, that disagreement is pointless; should liberals and republicans agree on the usage of interference, properly understood, that disagreement would disappear. And one could say that republicans diagnose the disagreement over whether or not unfreedom requires interference as legislatively verbal: since relevant counter-examples of unfreedom do not involve liberals' understanding of interference, that disagreement has a point; should liberals and republicans agree on the usage of liberals' understanding of interference, that disagreement would persist precisely because its point is legislating the right way of using the concept of unfreedom, of which interference is not a necessary condition.

Which diagnosis is correct? In order to address this question, I will bar the contentious terms employed in the disagreement by reconstructing it with the terminology presented in Part II.

### 1.1 Does Unfreedom Require Interference?

In Part II, I described three domains around which limitations underpinning ascriptions of unfreedom tend to cluster. One of these domains concerned the objects of practical activity, or option sets: I described how an option set could be limited by restraints rendering at least one option either unavailable or ineligible, all else being equal. I then defined an option as any possible course of action, and the other relevant notions in the following way:

Unavailability: an option X is unavailable iff, were an agent willing to perform X, he would not be capable of performing X.

Ineligibility: an option X is more ineligible iff, were an agent willing to perform X, the present or future costs of him performing X are higher with R than would be without R, all else being equal.

Moreover, I identified the degree of ineligibility resulting from restraints imposing present costs with the least cost the agent would incur in overcoming R, and the degree of ineligibility resulting from restraints imposing future costs with the least cost the agent would incur in overcoming R multiplied by the objective probability of R's occurring.

Importantly, these notions were made insensitive to 'internal' limitations (i.e. limitations on the formation and exercise of the will) by their counterfactual gloss: agents'

practical attitudes and beliefs are irrelevant for describing limitations on option sets, since their description only takes into account the counterfactual worlds where agents are willing to perform X, and the costs associated with this performance.

Finally, it should be noted that these notions can be extended to entire option sets. Thus, regarding option sets, ordinary distinctions between unfree actions, choices, and persons fail to register relevant theoretical differences. That is, if actions, choices, and persons are ascribed unfreedom on the basis of limitations on option sets, their differences can be accounted for quantitatively rather than qualitatively: an agent's action can concern a single option, an agent's choice concerns at least two options, and an agent's person can concern many option sets (which can be constituted of other option sets, and so on). This point is important because some theorists portray those distinctions as having theoretical relevance. However, regarding option sets, they can be accounted for in the same theoretical terms (viz. unavailability and ineligibility), and their intuitive differences registered as quantitative rather than qualitative. Therefore, when theorists portray those distinctions as having theoretical relevance, they might be concerned either with other kinds of limitations on practical activity, or with the significance of particular options or option sets. Both are valid concerns, but orthogonal to accounts of limitations on option sets.

Equipped with these notions, consider the following description of a situation between two agents, A and B, that extends from time t to time t':

At time t, agent A's option set does not include any option affecting B's option set.

At time t, agent B's option set includes option X, which is both available and eligible to B.

At time t', agent A's option set includes at least one of the following options: (i) preventing B from being capable of performing X, (ii) imposing present or future costs to B's performance of X. Call the subset composed of (i), (ii), or both (i) and (ii) subset D. Moreover, there are no restraints imposing present or future costs to A's performance of D. Therefore, subset D means that rendering X either unavailable or more ineligible to B is both available and eligible to A.

Necessarily, the passage from t to t' entails two changes in B's option set. First, rather than including option X, it now includes option X *conditional on A's performance of D* - henceforth simply option X'. In other words, option X became unavailable to B by being

replaced by option X'. Second, at any given time[63], whether or not X' becomes unavailable or more ineligible to B is exclusively determined by whether or not A performs D, all else being equal. In other words, the probability of X' becoming unavailable or more ineligible to B is identical to the probability of A performing D, all else being equal.

In short, necessarily, the passage from t to t' involves an expansion of A's option set that entails a limitation on B's option set, all else being equal.

The situation just described depicts what happens to the option sets of agents that establish a relation that could be called domination; in this case, A dominates B regarding option X'. Assuming this situation to persist indefinitely, all else being equal, one could say that A and B stand in a relation of pure domination regarding option X': since B's access to X' is indefinitely exclusively determined by A, one could say that A purely dominates B regarding option X'.

In order to make the description above more concrete, consider the Leash Case. At time t, agents A and B are nor related in any relevant way. Routinely, after lunch, B goes for a walk in a nearby park; this is an almost costless option that B ordinarily has. At time t', A buys B's dearest park, and puts in place a weird regulation: only when attached to a retractable leash handled by A, who stays in a protected cabin, can users walk in the park. B, a creature of habit, is informed that the use of the leash is mostly inconsequential, apart from the inconvenience of wearing it, and accepts using it in order to avoid the greater inconvenience of habit change. Thus, from t' onward B's walks in the park were replaced by walks in the park attached to a leash handled by A. However, there is a catch: since A wants users to avoid the east side of the park, but does not want to retract the leash, whenever a person comes close to that area, A uses the leash to deliver an electric shock. Thus, from t' onward B's walks in the east side of the park became more costly.

In other words, the passage from t to t' is marked by the establishment of a relationship between A and B that necessarily expands A's option set to include the intrusive options provided by the regulation, and necessarily limits B's option set by rendering unavailable the option to walk in the park without a leash handled by A. Moreover, the new relationship also contingently limits B's option set by rendering the new option amenable to being curtailed by retractions of the leash and/or burdened with the cost of an electric shock, which is almost certain to occur on the east side of the park.

---

[63] That is, at any given time while A's option set includes D.

The description above, and the outlandish Leash Case used to illustrate it, help elucidate the disagreement over whether or not unfreedom requires interference. In this disagreement, republicans emphasize what happens to option sets in the passage from t to t'; more specifically, they emphasize that B's option set at t' is more limited than B's option set at t, since it necessarily involves X being rendered unavailable by being replaced by X' (i.e. X conditional on A's performance of D). On the other hand, liberals emphasize what happens to option sets after t'; more specifically, they emphasize that B's option set at t' is more likely to be further limited, since it contingently involves X' being rendered unavailable or more ineligible, depending on whether or not A performs D.

Therefore, regarding option sets, the notions of unavailability and ineligibility suffice to account for the limiting effects of relations of domination. Consequently, if the concept of interference is analyzed in terms of unavailability and ineligibility, then it can account for the limiting effects of relations of domination on option sets. Therefore, if liberals analyze the concept of interference in these terms, as they seem to do, then they can account for the limiting effects of relations of domination on option sets.

If republicans do not recognize this[64], it is probably because they tend to conflate effects on agents' option sets with effects on agents' wills, as is suggested by their tendency to explain relations of domination in terms of the phenomenology that usually marks the passage from t to t'. Usually, the passage from t to t' is marked by A's invigilating attitude toward B's performance of X', and by B's obsequious or furtive attitude toward A with respect to the performance of X'. These attitudes are typical indicators that A and B stand in a relation of domination: A's invigilating attitude typically indicates the fact that A enjoys option D, whereas B's obsequious or furtive attitude typically indicates the fact that B enjoys option X', which is a more precarious option than the one that obtained at t. Albeit important, and probably central to any story about how domination affects agents' wills, these typical attitudes are manifestations of what happens to option sets in the passage from t to t', without themselves constituting limitations on option sets.

---

[64] In fact, republicans often do recognize this: 'the most salient explanation of why coercion and similar initiatives affect the freedom of a choice [is that,] while unchecked coercion does not simply remove any of the options by which a choice is characterized, it does replace one or another option' (Pettit 2008b, p. 120. See also Pettit 2008a, p. 215; Skinner 2008, p. 96). However, this curious concession is not accompanied by its necessary conclusion, namely, that unfreedom requires interference.

On the other hand, if liberals do not recognize this[65], it is probably because they tend to overlook that the passage from t to t' necessarily involves a limitation on B's option set by rendering X unavailable. Liberals correctly point out that A's enjoyment of D, or A's power to interfere arbitrarily, typically increases the likelihood of X' being rendered unavailable or more ineligible to B, that is, of B's being interfered with. However, they tend to overlook that, in the passage from t to t', A has already interfered with B by rendering X unavailable, and continues to do so as long as D remains available and eligible. The Leash Case highlights this point: A can interfere with B by retracting the leash or by using it to deliver electric shocks, but these interferences are possible only because A has already interfered with B by attaching a leash to him, and continues to do so as long as the leash remains attached.

Finally, if republicans and liberals come to recognize that the above is entailed by an analysis of interference in terms of unavailability and ineligibility, then both must accept that the ALS indeed involves a non-sequitur, and is therefore unsound. However, *contra* liberals, the non-sequitur involved in the ALS does not rest on the fact that C. 1 and C. 2 do not follow from Pr. 1-3, properly understood; rather, it rests on the simpler fact that Pr. 3 is false: there cannot be slavery without interference. As a social institution, slavery presupposes pervasive and ongoing interference with a slave's option sets: entire option sets concerning his life, work, family, and community are rendered unavailable to him *even if he has no master* (e.g. if he is an escapee). There must exist a high degree of social coordination enabling a master to enjoy option D, and no prospect of social coordination in the opposite direction, for this situation to obtain. This is how masters come to acquire option D, or the power to interfere arbitrarily, or the entitlement to impose new restraints on slaves' option sets; the power to interfere is a *consequence* of extant and widespread interference. To put these ideas metaphorically: the fact that you are attached to someone else by a leash that is invisible or not used does not make it less interfering. In sum, and more generally, whenever A enjoys option D, B has already been interfered with: option X has been rendered unavailable by being replaced by option X'.

I conclude that the liberal diagnosis of the disagreement over whether or not unfreedom requires interference is correct, but generally for the wrong reasons: this

---

[65] Goldwater 2020 is a notable exception that makes exactly the same point: 'Just as being in a jail cell entails unfreedom due to the jail's actual interference with one's range of choices, so too does being a slave entail unfreedom due to slavery's actual interference with one's range of choices. The constraints or obstacles that are jails and slavery are not merely empirically correlated with unfreedom' (p. 150). Lang 2012 points in the same direction.

disagreement is merely verbal, not because republicans' putative counter-examples of unfreedom contingently involve interference, but because, regarding option sets, such counter-examples necessarily involve interference.

## 1.2. Does Unfreedom Only Require Interference?

But republicans might argue that the original disagreement over whether or not unfreedom requires interference conceals a more important disagreement over the following liberal claim:

Unfreedom only requires interference[66].

Republicans, who deny this claim, might argue that it is this disagreement that reveals how their account of unfreedom differs from the liberal account, and in a way that recommends the former against the latter. By denying that interference is a sufficient condition for unfreedom, republicans might insist that it is not interference simpliciter that is unfreedom-realizing, but a subset of it; and that it is the basis upon which this subset is selected that should be of primary concern in any account of unfreedom. In other words, republicans might insist that what makes it the case that an instance of interference is unfreedom-realizing is something else, or that unfreedom is a property that supervenes on interference and something else. This something else, of course, is the reason why they conceptualize unfreedom in terms of, say, domination. Thus, rather than being understood as denying that unfreedom requires interference, republicans might be understood as trying to articulate an account of unfreedom that explains why interference can be unfreedom-realizing.

In order to elucidate the disagreement over whether or not unfreedom only requires interference, I will once again bar the terminology employed by liberals and republicans and reconstruct it with the terminology presented in Part II.

In Part II, I described three domains around which limitations underpinning ascriptions of unfreedom tend to cluster. One of these domains concerned the formation of the will: I described how practical activity could be limited in its formation, mainly as a result of limitations impinging on the formation or on the origins of practical attitudes. The former kind of limitations I labeled irrationality, and the latter, inauthenticity. I pointed out that both kinds ordinarily underpin ascriptions of heteronomy, because both convey the sense that an

---

[66] Strictly speaking, liberals endorse a version of this claim that is qualified by the background agreement mentioned earlier.

agent's will is not self-ruled. Moreover, I broke down inauthenticity in two types, vicariousness and dissociativeness. In order to elucidate the republican position on the disagreement on hand, one needs to focus on those limitations that render practical activity vicariously inauthentic.

Practical activity is vicariously inauthentic iff it is externally determined. When an agent's practical attitudes (desires, practical judgments, or intentions) have an alien origin, his practical activity can be said to be externally determined, and therefore vicarious. If the agent has the capacity to be the source of his own practical attitudes, vicariousness is often considered a limitation on his practical activity, since he is no longer its author; he is, as it were, the actor of, or a proxy for, another person's will. Familiar examples of vicariousness and of its distinctive limiting effects include cases of successful hypnosis, manipulation, persuasion, and indoctrination: agents whose practical activity is externally determined in these ways are usually considered defective because they are limited in their capacity of being the source of their own practical attitudes.

Notice that vicariousness need not involve any kind of limitation on option sets: actual profiles of availability and eligibility can remain the same despite actual vicarious changes in practical attitudes. However, it is often the case that the point in modifying actual or believed profiles of availability and eligibility of options, especially in a limiting way, is to explore the typical effects that such modified profiles have on the formation of agents' wills. More specifically, by rendering options unavailable or more ineligible, one can produce typical impeding and deterring effects on agents: not only will such options become impossible or more difficult to perform, they will also be less likely to become the object of agents' practical attitudes. Similarly, by leading agents to believe that options have been rendered unavailable or more ineligible, one can produce typical impeding and deterring effects on agents: whether or not such options have become impossible or more difficult to perform, they will be less likely to become the object of agents' practical attitudes. In other words, actual or believed limitations on option sets can be instrumental to the production of limitations on the formation of the will, including vicariousness.

Vicariousness that is indirectly obtained by means of modifications on actual or believed limitations on option sets seems to constitute the core of the republican account of unfreedom. At least, this is what is suggested by the way they conceptualize unfreedom, their focus on the phenomenology of domination, and their preferred cases of unfreedom.

First, republicans have variously conceptualized unfreedom in terms of power to interfere arbitrarily, unfitness to be held responsible, uncongenial control, uncontrolled control, alien control, dependence, or domination. These various ways of conceptualizing unfreedom are intended to convey the idea that interference is unfreedom-realizing only insofar as the interferee does not control it. More specifically, these notions have been fleshed out in terms of two requirements: the interference must occur with impunity, and the interference must not track the interferee's avowed interests. The first requirement registers a feature of the interferer's option set, viz., that his option set at t' is not rendered ineligible. The second requirement registers the intuition that only unauthorized interference is unfreedom-realizing. In other words, only interference that is not somehow the expression of the interferee's will can be unfreedom-realizing. One could plausibly reconstruct this requirement as follows: the interference must infringe the interferee's capacity to be the source of his own practical attitudes. Republicans therefore circumscribe the subset of unfreedom-realizing interference in ways that are highly suggestive of vicariousness.

Second, and as pointed out in the previous section, republicans' focus on the phenomenology of domination suggests that they are primarily concerned with effects on agents' wills rather than on option sets. That is, their focus on attitudes of invigilation, obsequiousness, furtiveness; on the hindering effects of misinformation on deliberation[67]; on the anxiety resulting from the uncertainty of perceived domination[68]; on the incapacity to look a perceived dominator in the eye, and on other typical markers of relations of domination that do not themselves constitute limitations on option sets, suggests that their central concern is with vicarious limitations on the formation of the will[69].

Third, republicans preferred cases of unfreedom tellingly involve vicariousness. Besides benevolent master cases, they also admit that hypnosis and brainwashing can be unfreedom-realizing[70]. Moreover, republicans usually illustrate interference that is not unfreedom-realizing with cases that are accounted for in terms of non-vicariousness:

---

[67] Pettit 2008b, pp. 110-1; 2015a, p. 383.

[68] Skinner 2008, p. 90

[69] Despite being a typical marker of relations of domination, this complex phenomenology can be rooted in other causes: for example, it can be rooted in a person's timorous temperament (perhaps the person scores high on traits neuroticism and agreeableness), in delusion, or even in respect. Alternatively, this phenomenology may not accompany relations of domination at all, such as when a person is unaware of being dominated (see also Lang 2012, pp. 288-90). Republicans recognize this, but insist in using this phenomenology as a heuristic guide as to 'whether there is a reason why people – even non-timid, non-cowardly people – should live in the fear and deference that a power of interference might inspire. And that depends on their social situation in relation to others, not on how they feel about it' (Pettit, 2015, p. 693).

[70] Pettit 2008b, p. 110.

When Ulysses was held to the mast by his sailors [...] he was not dominated by their interference, because they were acting only on terms that he laid down. They were not imposing an alien will in the practice of such interference but merely channeling his own will, as that had been expressed in his instructions to them (Pettit, 2016b, p. 189).

Finally, that vicariousness is at the core of the republican account of unfreedom is what transpires from statements like the following: 'Slaves whose choices happen never to fall out of conformity with the will of their masters [...] nevertheless remain wholly bereft of liberty. They remain subject to the will of their masters, unable to act according to their own independent will at any time' (Skinner, 2008, pp. 89-90); 'Did you depend on getting the leave or permission of another in order to make a choice, then that person's will would be in ultimate charge of what you do, not your own' (Pettit, 2016a, p. 7). More thoroughly:

You will be subject to my will in choosing between the options just to the extent that whether you can choose as you wish depends on the state of my will. And the fact that I have a power of interference means that it does depend on the state of my will. Let me be good-willed toward you, letting you have your way, and you can choose as you will. But let me turn nasty and you will not be able to choose as you will. Thus, you depend on the state of my will for the ability to choose according to your will. If you can choose according to your will, then that is only because I am happy to let you do so; you enjoy my grace and favor. The important point is that I am the one ultimately in control. What you can do you can only do because I let you do it. (Pettit, 2015a, p. 385)

Thus, republicans' denial of the claim that unfreedom only requires interference should be understood in light of their positive account of unfreedom. According to this account, limitations on option sets are subsidiary to limitations on the formation of the will: it is not interference simpliciter that is unfreedom-realizing, but the subset of it that renders an agent's will vicariously limited. In other words, the republican account of unfreedom gives explanatory priority to vicariousness: it is vicariousness that makes it the case that an instance of interference is unfreedom-realizing. Therefore, and more precisely, republicans deny the claim that unfreedom only requires interference because they endorse the rival claim that unfreedom only requires interference, or believed interference, that vicariously limits the formation of the will.

Should liberals accept this vicariousness-based account of unfreedom? Carter claims that their 'negative definition prescinds from questions about the nature of the will and the conditions for its autonomy or freedom' (2008, p. 73). However, if liberals admit that intimidation, misrepresentation, indoctrination, and other limitations that need not involve interference, can be unfreedom-realizing, or if they admit that the unfreedom involved in, say, the Orwellian nightmare of propaganda and surveillance cannot be entirely accounted for in terms of limitations on option sets, their story is incomplete. Whether or not republicans' account provides the correct way of completing that story is a further issue that is outside the scope of the present discussion[71].

I conclude that the disagreement over whether or not unfreedom only requires interference is legislatively verbal, because its point is to legislate a role to the concept of unfreedom that reflects an account of unfreedom centered around vicarious limitations on the formation of the will. That is, its point is to legislate interference, or believed interference, that vicariously limits the formation of the will as a sufficient condition for applying the concept of unfreedom[72].

---

[71] Personally, I think neither liberals nor republicans succeed in explaining why the limitations upon which they concentrate can be unfreedom-realizing, because they do not provide any explicit conception of individual potential in agency that help circumscribe which limitations on practical activity can be unfreedom-realizing. However, in this Part I am merely interested in illustrating a method for handling verbal disagreements, rather than in assessing the theoretical adequacy of their accounts.

[72] Strictly speaking, if *believed* interference that vicariously limits the formation of the will is sufficient for unfreedom, then unfreedom does not require interference. However, I think republicans would like to ground believed interference in actual domination, which requires interference, thereby once again repelling that conclusion. I leave such complexities to republicans.

**2. Normative Disagreements Over Unfreedom.**

In part I, I distinguished between two types of normative disagreement:

(i) disagreement concerning whether or not an item satisfies the role property of either realizing or promoting a given hierarchically ordered possible world;

(ii) disagreement concerning the correct way of specifying a given hierarchy of possible worlds.

I said that type (i) disagreement is straightforwardly substantive: parties assert incompatible contents because they disagree over how the world is normatively. I also said that type (ii) disagreement is prone to give rise to verbal disagreements, especially of the legislative kind: parties disagree over the conditions for something to count as good, bad, right, or wrong, which in turn can lead them to disagree over which role to ascribe to terms whose usage is guided by normative judgments.

I will now present, in the form of a dialogue, some type (ii) normative disagreements that might underlie verbal disagreements over unfreedom in the political context.

A: I think that disagreements over how to classify things as instances of unfreedom are sometimes rooted in deeper disagreements concerning the disvalue, or the badness or wrongness, of unfreedom. In such cases, a more promising strategy is to move from the linguistic surface to the normative depth, so to speak.

B: I agree. How do you propose we proceed?

A: We could begin with an agreement on the following normative judgment: political tyranny is bad. Since, intuitively, political tyranny necessarily involves unfreedom, we could proceed by investigating the disvalue of the former in order to gain insight into the disvalue of the latter. This should give us a lateral entry into political theorists' verbal disagreements over unfreedom.

B: I cannot refuse a reasonable proposal. Indeed, I agree that political tyranny is generally bad, and for the following reasons: it hinders error correction and predictability; it undermines trust, thereby hindering voluntary cooperation and competition; it generates insecurity and sometimes severe instability; finally, it blocks individuals' ability to master their own lives.

A: The features just listed would certainly figure in any plausible account of tyranny's badness. But perhaps we should concentrate on those features that directly bear on individuals, since they are the primary locus of unfreedom.

B: Sure. Two features stand out: under tyranny, most individuals have no say in collectively binding decisions, and most individuals' options are more likely to be obstructed. Since those in power are unrivaled in their capacity to unilaterally impose their will, most individuals will be pressured into conformity by means of a perverse incentives structure. Typically, this structure will reward individuals who comply, lie, and spy; however, its most distinctive characteristics are punishment and threat of punishment. That is, the primary tool of tyranny is coercion: it is through force, the threat of its use, and the dissemination of fear that those in power ultimately pressure most individuals into conformity. Living in an atmosphere of insecurity, most individuals' potential in agency will be severely curtailed, because both their wills and options will be severely limited. Hence their unfreedom and its accompanying badness.

A: Before unpacking this story linking tyranny, unfreedom, and their accompanying badness, I would like to inquire about one apparent implication of it. I noticed that the story does not discriminate between ways of life, and that the unfreedom accompanying tyranny applies across the board. I mention this because, on some accounts, unfreedom resides not in restrictions simpliciter, but only in those that also limit individuals' flourishing, that is, individuals' propensities and capacities that constitute a superior way of life. Apparently, our story implies the rejection of such accounts. For instance, is it compatible with licentiousness?

B: Yes, that story is compatible with licentiousness; the only requirement is that one's individual's options do not impair another's. Remember, our subject is unfreedom in the political context. In this context, historical experience has convincingly shown that, in the trade-off between licentiousness and state sponsored flourishing, the badness resulting from the latter tends to greatly surpass the badness resulting from the former. That is, allowing for inferior ways of life is the price of preventing widespread state intervention. In any case, it is now generally agreed that states' role does not include the perfecting of individuals; at best, it includes the provision of some of the preconditions for individuals' development. Our story reflects this position by not discriminating between ways of life in accounting for the unfreedom involved in political tyranny.

A: I think the trade-off you mentioned merits further analysis, but we should move on to more consequential issues. Our story linking tyranny, unfreedom, and their accompanying badness strikes me as rather securitarian. By highlighting coercion as the primary tool for influencing individuals' wills and options, our story suggests that we disvalue unfreedom because it involves insecurity. Now, compare this with the reasons why the ancients valued the opposite state, that is, the state of liberty. Liberty designated the state of individuals capable of actual autonomy, understood as the actual capacity to provide, protect, and promise. Such individuals were assigned a special social status because they were endowed with this rare capacity; that is, their social status was a privilege conceded in recognition of that capacity. Presumably, in order to achieve such a state, those individuals must have been less risk-averse than the regular individual, and they must have been willing to bear the consequences for their actions. This is exemplified by the old adage: *coactus voluit, attamen voluit*, which roughly means 'wished coerced, nonetheless wished'. For modern minds accustomed with the view that duress and misrepresentation vitiate contracts, that adage seems utterly cynical; however, for those individuals, it was a direct expression of their singular capacity: if they were fooled or overpowered, this would be a reason for shame more than for resentment; it would be a sign of a fissure in their autonomy. Similarly, if they were assisted by more powerful agents in bearing the untoward consequences of their actions, this would be seen as a recognition of impotence, of a decrease in their singular capacity, rather than a restoration of independence. In short, they valued their state of liberty out of pride in their exceptional capacities, not out of fear of others' capacities. Conversely, one could expect them to disvalue unfreedom not because it involved insecurity, but because it testified to an individual's inferior capacity for autonomy.

B: I am not sure if I understand the point you are trying to make with this speculative history of the disvalue assigned to unfreedom and the sentiments behind it.

A: My point is the following: if unfreedom is disvalued because it involves insecurity, then we should disvalue liberty, because it also involves insecurity. Moreover, it is a well-known irony of nature that worrying about insecurity is a self-fulfilling prophecy, that is, that the search for security tends to lead to insecurity. At the psychological level, the search for security tends to hinder the development of capacities required to deal with risks, which tends to result in overall weakness, which tends to increase vulnerability to risks. At the social level, the search for security tends to trigger security dilemmas, which increase risks; it also tends to lower social risk-tolerance curves, which tends to trigger precautionary behaviors that

end up increasing risks. Therefore, if the disvalue of unfreedom resides in insecurity, it should be tackled cautiously in order to avoid prompting these dynamics.

B: All this is very well said, but somewhat besides the point. Indeed, insecurity can be a major source of unfreedom - after all, force and the threat of its use are paramount among the means to block access to options and to adjust agents' wills. However, it is only a subset of sources of unfreedom: options can be rendered unavailable or ineligible without being attached to costs involving harm to valued things, and the will can be adjusted by various other means. Therefore, the disvalue of unfreedom does not reside in insecurity, although it can be, and often is, aggravated by it.

A: I appreciate the clarification regarding the narrow overlap between unfreedom and insecurity, but I am afraid our discussion of the disvalue of tyranny and unfreedom is being too dismissive of the quality of the individuals that constitute a polity. After all, the popular prejudice stating that the majority of people are not animated by a desire to dominate is at the very least questionable; the spirit of tyranny can possess an entire body politic as well as individuals in positions of power. In fact, as that great account of tyranny which is the biblical narrative of Exodus shows, tyranny tends to be longed for as soon as unknown risks emerge; and one can expect the situation to be even worse in polities providing people with the occasion to participate in tyranny. Since, in the final analysis, the only bulwarks against tyranny are the aristoi, with their jealous self-reliance, and, ultimately, individuals' courage, one could even say that their absence is a necessary condition for political unfreedom. The point I am trying to make is that, in order to prevent unfreedom, the body politic needs a cultural soul of a certain quality –.

B: Let me interrupt you there, because I thought we had already established that our story is compatible with licentiousness. I sympathize with your bewilderment; if you will, we could discuss the trade-offs that led to that compromise. However, this would not lead us very far in understanding the normative disagreements that can be at the root of the conceptual disagreements between liberals and republicans.

A: How, then, should we understand their disagreement concerning the disvalue of unfreedom?

B: I think their disagreement is best understood from the perspective of the contrasting roles they assign to the state. Roughly speaking, republicans assign a more

demanding role to the state than liberals. Republicans' emphasis on vicariousness-based unfreedom allows them to do so; at the same time, it provides them with a justificatory theory of state authority. If state action is controlled by its citizens, then it does not render them unfree; moreover, it is justified, because it is expressive of their own will. In addition, the emphasis on vicariousness-based unfreedom provides the state with a goal: to reduce power asymmetries that render citizens' options vicariously limited. Liberals, on the other hand, are more interested in limiting state action than in justifying its authority or in providing it with a goal. Their emphasis on limitations on option sets, of whatever sort, allows them to do so: it establishes an internal constraint to any state interference, which requires extraordinary justification, preferably in terms of reduction of further interference. Liberals are generally more skeptical about the prospects of the inherent trade-offs involved in state interference being beneficial to its citizens. They could object, for example, that, since the state is the most powerful entity that has ever existed, the infinitesimal influence that citizens can have upon state interference offers just an illusion of control; and that, even if discursive control could be exercised, it would likely spiral out of control or be based on bad reasoning. Moreover, they could object that the use of state interference to prevent power asymmetries is likely to be counterproductive and undesirable: since any functional hierarchy produces power asymmetries, targeting the latter risks destroying the former.

A: In other words, republicans primarily take freedom to be an end of the state, whereas liberals primarily take freedom to be a buffer against it.

B: That's about right.

And so on.

The exchange above reveals some type (ii) normative disagreements that might underlie verbal disagreements over unfreedom in the political context. In particular, type (ii) normative disagreements concerning the state seem to underlie verbal disagreements between liberals and republicans: that is, their verbal disagreements seem to be rooted in their conflicting ways of specifying a hierarchy of possible ways of being a state. Since liberals seem to rank higher states whose actions are in principle more constrained, their emphasis on unfreedom as limitations on option sets allows them to assign a more constrained role to the state. And since republicans seem to rank higher states whose actions are in principle

justified, their emphasis on unfreedom as limitations on option sets that vicariously limit agents' wills allows them to assign a more spacious role to the state[73].

Finally, the exchange above left out the following important point of controversy:

Disagreements concerning hierarchies of ways of being an agent. Both liberals and republicans seem to endorse a normatively neutral conception of individual potential in agency as propensities and capacities. However, republicans' emphasis on vicariousness suggests that they hierarchize agents in terms of their capacity of being the source of their practical attitudes, ranking lower agents whose practical attitudes are externally determined. At least, this seems to be Philip Pettit's case: for him, the disvalue of unfreedom as domination lies in the fact that it undermines agents' capacity to have free will, which he conceives as 'nothing more or less than conversability' (2015a, p. 379), or 'deliberative control'[74]. When agents' options are conditional on another person's will, they cease to have deliberative control over those options. Thus, to the extent that domination entails vicariousness, it undermines an agent's free will, and is therefore bad/wrong. More precisely: domination is bad/wrong because it eliminates a precondition for the exercise of free will. In other words, Pettit specifies a hierarchy of ways of being an agent in terms of free will, and grounds the disvalue of unfreedom as domination in its undermining effects on free will.

Plausibly, some conceptual disagreements over unfreedom in the political context are inspired by these contentious normative views specifying hierarchies of ways of being a state and hierarchies of ways of being an agent. Debates over unfreedom could therefore benefit from making these conflicting normative views explicit: it would allow parties to directly address their primary source of conflict, and force them to articulate the reasons for espousing their implicit normative views, an endeavor that is tantamount to articulating a first-order normative theory. To the extent that disentangling these kinds of disagreements is a precondition for philosophical progress, normative disagreements should not be conflated with verbal disagreements.

---

[73] See Pettit 2015b for his views on this subject.
[74] See also Pettit 2008b, p. 220. See Pettit 2001 for a book-length account of freedom as fitness to be held responsible that suggests the existence of a justificatory normative system underpinning his conceptual work.

**CONCLUSION**

As I said in the Introduction, my goal in this dissertation was to provide methodological constraints that helped canalize a recurrent temptation in political philosophy, namely, that of pursuing victory for one's normative views by legislating on the use of language. I said that, when unconstrained, this temptation could be problematic because, to the extent that the truth about those views remained unexamined, it would involve philosophical cheating, and because, to the extent that the examination of those views was obstructed, it would make those who live by them their serfs.

My attempt to address this issue centered around the claim that normative disagreements should not be conflated with conceptual disagreements, and around a methodological procedure to distinguish between them. In Part I, I provided the basis for that claim and introduced that methodological procedure. Parts II and III were intended to illustrate the discussion developed in Part I: in Part II, I provided an analysis of the concept of unfreedom that was attentive to its context-sensitivity, a feature that is central to understanding how ascriptions of unfreedom can convey normative beliefs; and, in Part III, I applied the proposed methodological procedure to some disagreements over the concept of unfreedom between liberals and republicans in contemporary political theory.

My hope is that my methodological proposals help make normative disagreements more explicit, thereby advancing philosophical progress.

Am I a fool in thinking that temptation can be resisted with methodology? After all, the victory-seeking game has always tempted the truth-seeking game, under various disguises. Perhaps I am; nonetheless, one can at least find solace in the fact that, when one can see through a temptation's disguise, she will blush on account of her nakedness.

# BIBLIOGRAPHY

CARTER, I.; SHNAYDERMAN, R. (2019). The Impossibility of "Freedom as Independence." **Political Studies Review**, 17(2), 136–146.

CARTER, I. (2008). How are power and unfreedom related. In Cécile Laborde & John W. Maynor (eds.), **Republicanism and Political Theory**. Blackwell. pp. 58--82.

CEPOLLARO, B.; STOJANOVIC, I. (2016). Hybrid Evaluatives: In Defense of a Presuppositional Account, **Grazer Philosophische Studien**, 93(3): 458–88.

CÔTÉ, N. (2020). Weakness of Will and the Measurement of Freedom. **Ethics** 130 (3):384-414.

CÔTÉ, N. (2022). Measuring republican freedom. **Synthese** 200 (6).

DYKE, H. (2003). What moral realism can learn from the philosophy of time. In **Time and Ethics**: Essays at the Intersection. Kluwer Academic Publishers. pp. 11--25.

GOLDWATER, J. (2020). Freedom and Actual Interference. **Journal of Ethics and Social Philosophy** 17 (2).

GOODIN, R. E.; JACKSON, F. (2007). Freedom from Fear. **Philosophy & Public Affairs**, 35(3), 249–265.

JACKSON, F. (1998). **From Metaphysics to Ethics**: A Defence of Conceptual Analysis. Oxford, GB: Oxford University Press.

KRAMER, M. (2008). Liberty and domination. In Cécile Laborde & John W. Maynor (eds.), **Republicanism and Political Theory**. Blackwell. pp. 31--57.

LANG, G. (2012). Invigilating Republican Liberty. **The Philosophical Quarterly**, 62(247), 273–293.

LIST, C.; VALENTINI, L. (2016). Freedom as Independence. **Ethics** 126 (4):1043–1074.

LOVETT, F. & PETTIT, P. (2018). Preserving Republican Freedom: A Reply to Simpson. **Philosophy and Public Affairs** 46 (4):363-383.

PETTIT, P. (1996). Freedom as antipower. **Ethics** 106 (3):576-604.

PETTIT, P. (2001). **A Theory of Freedom**: From the Psychology to the Politics of Agency. Polity.

PETTIT, P. (2008a). Republican Freedom: Three Axioms, Four Theorems. In Cécile Laborde & John W. Maynor (eds.), **Republicanism and Political Theory**. Blackwell. pp. 102-130.

Pettit, Philip (2008b). Freedom and probability: A comment on Goodin and Jackson. **Philosophy and Public Affair**s 36 (2):206-220.

PETTIT, P. (2015a). Freedom: psychological, ethical, and political. **Critical Review of International Social and Political Philosophy** 18 (4):375-389.

PETTIT, P. (2015b). Freedom and the State: Nanny or Nightwatchman? **Public Health** 129 (8):1055-1060.

PETTIT, P. (2016a). Freedom and Other Robustly Demanding Goods. In Simon Derpmann & David Schweickart (eds.), **Philip Pettit**: Five Themes from his Work. Springer. pp. 3-16.

PETTIT, P. (2016b). Three Mistakes About Democracy. In Keith Breen & Allyn Fives (eds.), **Philosophy and Political Engagement**: Reflection in the Public Sphere. London: Palgrave. pp. 187-199.

PLUNKETT, D.; SUNDELL, T. (2013). Disagreement and the Semantics of Normative and Evaluative Terms. **Philosophers' Imprint** 13 (23):1-37.

RABERN, B. (2017). A Bridge from Semantic Value to Content. **Philosophical Topics** 45 (2):181-207.

ROBERTS, D. (2015). Review: Pekka Väyrynen, The Lewd, the Rude and the Nasty: A Study of Thick Concepts in Ethics. **Ethics** 125 (3):910-915,.

STOJANOVIC, I. (2016). Expressing Aesthetic Judgments in Context, **Inquiry**, 59(6): 663–85.

STREUMER, B. (2018). Reasons and Ability. In D. Star (Ed.), The Oxford handbook of reasons and normativity. Oxford: Oxford University Press.

STREUMER, B. (2017). **Unbelievable Errors**: An Error Theory About All Normative Judgments. Oxford, United Kingdom: Oxford University Press.

SWANTON, C. (1985). On the "essential contestedness" of political concepts. **Ethics** 95 (4):811-827.

SWANTON, C. (1992). **Freedom**: A Coherence Theory. Hackett.

VÄYRYNEN, P. (2013). **The Lewd, the Rude and the Nasty**: A Study of Thick Concepts in Ethics. New York, NY: Oxford University Press.

VÄYRYNEN, P. (2014). Essential Contestability and Evaluation. **Australasian Journal of Philosophy** 92 (3):471-488.

YALCIN, S. (2018) Expressivism by Force. In D. Fogal, D. Harris and M. Moss (eds.) **New Work on Speech Acts**. Oxford University Press, 400-429.

YALCIN, S. (2018). Belief as Question-Sensitive. **Philosophy and Phenomenological Research**. XCVII (1): 23-47.

YALCIN, S. (2021). Fragmented But Rational. In C. Borgoni, D. Kindermann, and A. Onofri (eds.) **The Fragmented Mind**. Oxford University Press. 156-179.

YALCIN, S. (2022). Context-Sensitivity by Way of Parameters. In E. Lepore and E. Sosa (eds.) **Oxford Studies in the Philosophy of Language**, Volume 2. 267-304.

YOUNG R. (1975). **Freedom, Responsibility and God**. Harper & Row, New York.