The Problem with Hart's Habits. Elements of a Humean Conception of Social Obligation

*Marco Segatti**

Abstract

In a few powerful pages in chapter 4 of *The Concept of Law*, Herbert Hart demolishes a "Habits and threats" conception of legal obligation, by posing three main challenges: the "efficacy challenge", the "compulsion challenge" and the "normativity challenge". This "Habits and threats" theory, says Hart, is an Austinian conception of legal obligation but Hart's arguments easily generalize to long-standing criticisms of a Humean conception of human agency and social motivation as well. The aim of this essay is to search for clues in Hume on ways to meet Hart's challenges and provide a conception of "habits" which doesn't fall prey to Hart's devastating critique. Through a somewhat manipulative interpretation of Hume's text, the essay proposes two tweaks to Hume's general framework and finds that they perform well on the efficacy problem, and on the compulsion problem. They perform only decently on the normativity problem, but I shall find good news in that too.

Keywords: Hart, Hume, Practice theory of rules, Habits, Social Obligation.

1. Introduction

In a few powerful pages in chapter 4 of *The Concept of Law* (CL, from hereinafter), Hart (1961) demolishes, what I shall call, a "Habits and threats" conception of legal obligation. In perhaps one of the most successful arguments in the contemporary history of legal philosophy, Hart claims that habits and threats are not enough to account for social rules by posing three main challenges to a "Habits and

* Istituto Tarello per la Filosofia del diritto, Dipartimento di Giurisprudenza, Università degli Studi di Genova, via Balbi 30/18, 16126 – Genova, Italia, segattimr@gmailcom.

Part of the research has been supported by the European Union, within the framework of the HORIZON WIDERA program, project no. 101079177 titled "Advancing Cooperation on the Foundations of Law." The project is carried out by a consortium consisting of the Universities of Belgrade, Genoa, Lisbon, and Surrey. The unit at the University of Genoa has been led by Prof. Andrej Kristan.

ANALISI E DIRITTO 1/2025: 177-197
doi: 10.4454/qj1a7g51

threats" theory. First, it cannot distinguish between merely socially convergent, and repetitive behavior on one side, and behavior compliant with a social rule on the other (call this the "efficacy challenge"). Second, it reduces the internal aspect of an obligation to a mere feeling of compulsion (call this the "compulsion challenge"); consequently, and third, it cannot account for the normative force of a social rule (call this the "normativity challenge").

This "Habits and threats" theory, says Hart, is an Austinian conception of legal obligation[1]. But its main building blocks are unmistakably Humean. Indeed, Hart's arguments in chapter 4 easily generalize to long-standing criticisms of a Humean conception of human agency and social motivation. The aim of this essay is to go against currents, however, as it searches for clues in Hume's *Treatise*[2] on ways to meet Hart's challenges and provide a conception of "habits" which doesn't fall prey to Hart's devastating critique[3]. Through a somewhat manipulative interpretation of Hume's text, the essay proposes two tweaks to Hume's general framework and show how they can be used to push forward the dialogue with Hart. Indeed, the paper argues that the two tweaks have something helpful to say on all three challenges[4].

To set the stage, the second section provides a summary of Hart's arguments and traces their significance in the recent history of legal philosophy. As we shall see, it is difficult to overestimate it: Hart's argument in chapter 4 is one of the most successful philosophical arguments on the law in the 20th Century, to say the least. I begin by drawing a neglected connection with Hart's arguments in chapter 7 of CL against predictive theories of law and legal obligation and then show how the arguments in chapter 4 complement these later (in Hart's presentation) arguments and strike out a few good replies by predictive theories of legal obligation.

Section 3 moves to Hume and shows how Hart's conception of habits fits well with a canonical interpretation of a Humean theory of human agency and motivation. Indeed, Hart's criticism to Austin in chapter 4 easily generalizes to this canonical interpretation. Arguably, Hume's *Treatise* gets into trouble when it takes "pleasure" to be the sole objective of choice: Hume's conception of human conduct significantly reduces the variety of goals and motivations, which human behavior displays.

Habits help somewhat because they rationalize the most visible failures of this hedonist interpretation. Perhaps we do not always directly pursue pleasure (or

---

1 See Austin 1831.

2 See Hume 1978: referred to by volume, part, and section.

3 That there is a "problem" with Hart's characterization of habits, has not, however, gone completely unnoticed in contemporary philosophy of law. See, for example, MacCormick 2007: 61 (I owe this reference to Bruno Assalin). See also Krygier 1982 and Moles 1987.

4 My own formulation of the two tweaks heavily borrows from Dewey (1922). Since the exact relation between Hume and Dewey's respective conception of human agency and motivation is not at all an easy interpretational matter, I won't further elaborate on these pragmatist influences here. Luca Malagoli has greatly helped me to navigate these pragmatist influences.

avoid pain) on *every* occasion of choice, but we become habituated to ways of pursuing it (or avoiding pain), thus either repeating these sets of dispositions in the conscious and rational hope for similar end-results or thoughtlessly making them cheaper routines.

But this is where chapter 4 comes in: human social commitments are also more than mere smart repetitive training, or thoughtless routines, because they set up standards for social criticism, and they are not merely socially convergent behavior. Also, social rules are more than mere compulsions because they involve an internal, normative, aspect, which defines one's attachment to them.

Section 4 proposes two tweaks to Hume's framework and shows that they are quite successful at keeping a large chunk of Hume's insights, while blocking forceful criticism. Here they are:

1) Hume's pleasure is not the sole object of choice, a "like – experience", an experience at the terminal point of activity, and which motivates it as a reward; Pleasure is a "want – experience", an experience at the moment of choice, and which pushes it forward. Whenever Hume writes pleasure, read stimulation.
2) Habits are a bundle of impulses, expectations, abilities, and environmental variables (including one's social capital), which reduce the relative marginal cost of cognitive awareness and motivational control over one activity or other. Formally, they define intertemporal complementarities between activities in the past, present, and future. Whenever Hume writes "original passion", read impulses. And whenever Hume writes "secondary indirect passion", read this definition of "habits".

Section 5 articulates these two tweaks and the resulting conception of human conduct in the contemporary jargon of standard microeconomics. The result shouldn't be confused with a philologically accurate representation of Hume's classic conception of human social and moral psychology, however. Rather, it is meant to show how relatively small changes within the latter affect large gains in its explanatory power.

Section 6 ties all these different threads together and test the two tweaks against Hart's three challenges. It finds that the tweaks perform well on the efficacy problem, and on the compulsion problem. They perform only decently on the normativity problem, but I shall find good news in that too.

2. Hart's Three Challenges to a "Habits and Threats" Theory of Social Obligation

Hart begins his argument with a famous thought-experiment (See CL, chapter 4, section 1):

> We shall suppose that there is a population living in a territory in which an absolute monarch (Rex) reigns for a very long time: he controls his people by general orders backed by threats requiring them to do various things which they would not otherwise do, and to abstain from doing things which they would otherwise do; though there was trouble in the early years of the reign, things have long since settled down and people can be relied on to obey him.
>
> […] It should be noted that, on this account of the social situation under Rex, the habit of obedience is a personal relationship between each subject and Rex: each regularly does what Rex orders him, among others, to do. If we speak of the *population* as "having such a habit", this, like the assertion that people habitually frequent the tavern on Saturday nights, will mean only that the habits of most of the people are convergent: they each habitually obey Rex, just as they might each habitually go to the tavern on Saturday night.

This is a very simple society, says Hart. But it offers a useful mental experiment because it fits with the idea of putting "habits of obedience" at the very foundation of a social order. Would we say that this society had a legal system? Or are "habits" enough to account for social rules, and especially those social rules, which confer obligations (legal or not)? Hart says no and believes that this refutation is instructive to identify what's missing in a "Habits and threats" theory of legal obligation.

Now, there are certainly several ways of reconstructing Hart's arguments here[5]. For my purposes, it is helpful to begin by drawing a few connections between Hart's arguments in chapter 4 and his later arguments in chapter 7 against predictive theories of legal obligation.

Recall Hart's definition in chapter 7 of CL. A predictive theory of legal obligation is one according to which the statement "I have an obligation to do X" means "The judge will sanction me with S, if I do Y instead of X". One of Hart's arguments in chapter 7 is that this predictive (and verificationist) conception of legal obligation doesn't reflect ordinary meaning and use at all. We often say we have an obligation even when we can't (or before we can) reasonably expect a sanction in case of non-compliance. That is, this predictive (and verificationist) definition collapses the existence of a legal obligation (circumstances C_1 and C_2 which make

[5] Compare, for example, Dworkin's (1977) arguments in *The Model of Rules 2*, Postema's (1982) (humean) account of conventions, and even Leiter's (2007) realist arguments against predictive accounts of the *concept* of social and legal obligation. Compare also Postema 1982; Coleman 2001 and 1982; Shapiro 2006, 2005 and 2002 for the development of a, so-called, *practice-conception* of social rules. On the reception of Lewis' (1969) Humean theory of conventions by the philosophical literature on social rules, see also Gilbert 1990, Green 1999, Dickson 2007, Coleman 2001 and Celano 2016 on the skeptical side, and Marmor 1996, 2007, 2009, Guala, Hindriks 2015, Guala 2016, and Bicchieri 2006, adding different refinements to the original Lewisian account. Even a cursory account of this large literature goes well beyond the limits of this paper. If my arguments here go through, however, then we have reason to think that the conception of habits which is provided in the present article can be useful for understanding the conventional nature of law and the nature of social practices generally.

"I have an obligation to do X" true), and one aspect of its efficacy (the fact that the justice-system does a fairly good job in responding to violations with predictable legal consequences).

Add habits, even in the limited sense of thoughtless repetition of past activity, and Hart's pointer in chapter 7 of CL loses some of its bite. Sure, we can meaningfully say "I have an obligation to do X, but I am not going to do it since no one will catch me". Or sometimes we comply with a social rule even when we don't know, or don't care, if someone will catch us if we don't. But habits can square these linguistic phenomena with a predictive theory with relative ease. Sometimes we comply with a social rule (or we don't) because we have become habituated to compliance by threats and actual sanctions (to ourselves and to others). These habits are not themselves necessarily thoughtless, or unintelligent, because they are often the result of social adjustments to stable circumstances. But once habituated, we follow through thoughtlessly, and without always adjusting our behavior to any new piece of information on expected sanctions (indeed, perhaps even either partially or completely unaware of what we are doing and of why on earth we keep doing it).

So, a "Habits and threats" theory can account for the existence of a social obligation (e.g.: "in social group A, there is a shared social habit of doing X in C_1 and C_2" and "This social habit was formed by habituation to threats and actual punishments") and distinguish it from its judicial efficacy ("the justice-system predictably punishes most wrong-doings according to the threats issued"). All the while, it can still consider itself a predictive (and verificationist) theory. Knowing the law (or social rules more generally) is still just about being able to make smart bets on observable social behavior; but "I have an obligation to do X in C_1 and C_2" means now "In my social group, there is a shared habit of doing X in C_1 and C_2", and "This habit was formed by habituation to threats and actual punishments" (which are, arguably, still statements expressed in predictive and observational language).

But this is where chapter 4 goes to work. Relatively thoughtless habits, compulsions and strategic behavior may well account for a lot, to be sure. But this is not even nearly enough, because we are still missing an account of the peculiar attitude of involved attachment of one subject taking up an internal perspective on a social rule. We are still failing, that is, the second part of the efficacy challenge (CL, 55):

> [G]eneral convergence or even identity of behavior is not enough to constitute the existence of a rule requiring that behavior: where there is such a rule, deviations are generally regarded as lapses or faults open to criticism.

Also, if we confuse "habits" with "social rules", we might mistake the internal aspect of rules with mere feelings of compulsion. Which means that we are failing the compulsion challenge too (CL, 57).

Finally, the normativity challenge: for something to be a rule and not a mere

habit, "not only is […] criticism in fact made, but deviation from the standard is generally accepted as a good reason for making it" (CL, 55, 57):

> What is necessary is that there should be a critical reflective attitude to certain patterns of behavior as a common standard, and that this should display itself in criticism (including self-criticism), demands for conformity, and in acknowledgements that such criticism and demands are justified, all of which find their characteristic expression in the normative terminology of 'ought', 'must', and 'should', 'right' and 'wrong'.

3. On Hume's Hedonist Psychology: Original, Secondary, Direct, and Direct Passions

The aim of this section is to present a canonical interpretation of Hume's central characterization of human agency and motivation in the *Treatise*. Then, the section shows that this Humean characterization falls prey to Hart's arguments in chapter 4 – that is, Hart's arguments in chapter 4 easily generalize and point to a structural deficiency of a Humean conception of human social behavior.

To fix ideas, let's begin with a few well-known distinctions and classifications[6]. Hume distinguishes between two general kinds of passions: original passions and secondary passions[7]. Original passions, also called implanted instincts, do not depend on prior experiences of pain and pleasure. They are incorporated within our constitutive make-up, assuring that, given perhaps common, or frequent, biological traits, we all find their satisfaction pleasurable. Examples of these original passions may include "hunger", "thirst", "sexual appetites", "empathy", etc. On average and at an adult age, humans are wired up in such a way that they perceive "hunger" with an empty stomach and find the ingestion of food pleasurable, just as they are wired up in such a way that they feel pain (and joy), whenever other people (ostensibly) feel pain (and joy) and they empathize with them.

Secondary passions depend on prior experiences of pleasure and pain instead. They are further differentiated between direct and indirect passions, depending on how these prior experiences of pleasure and pain have come to form the relevant passion[8]. Direct passions, like "desires" and "aversions", involve a single, direct, relation between prior experiences, and their cognitive content. A bee stung me a few months ago, and it caused pain. In response to it, I formed an idea of such pain, associating it with bees, which became the cognitive content of my aversion. I now

---

6 Here I closely follow Rawls' chapters on Hume in his *Lectures on Moral Philosophy* (see especially Rawls 2000: 37 ff.).

7 See Hume 1978: II, iii, 9

8 See Hume 1978: II, i, 2 – 5.

have a secondary, direct passion, which is an aversion to bees. In Hume's phrase, passions of this kind display a single, direct, relation between sensations and ideas. Whenever I see a bee, I anticipate the pain it shall inflict on me if it stings me, and this anticipation pushes me *now* to act in a way which, I believe, will allow me to avoid the pain.

Indirect passions (like "love", "pride", "humility", etc..) display a more complicated relation between prior experiences of pleasure and pain, and the cognitive content of the relative passion; in Hume's phrase, they display a double relation between sensations and ideas. Take Hume's analysis of pride, for example. Its cognitive content (i.e., whatever it is that we are inclined to think and focus on when we feel proud) is, Hume argues, "ourselves" – we act on "pride" whenever we act on an idea of ourselves, or a present or future characteristic of ourselves in relation to others more generally, which we find valuable or appealing and we want to achieve or protect it. But we form this "idea" and associate it with the gratification of pride by acting on another passion, and hence on another idea. We accomplish a goal by acting, say, on a desire (first relation between idea and "sensations": e.g., publishing my monograph will give me pleasure) and then, once this desire has been accomplished (the monograph is published) we redirect thought to ourselves and feel proud (second relation between idea and sensation).

Now, there is a rather canonical way of reading these classifications. Human conduct strictly depends on pleasure. Cognition plays a mediating role between past "like-experiences" (whatever people enjoyed or suffered in the past) and future ones: it projects means-ends regularities in past "like-experiences" (e.g.: "I feel pleasure whenever I eat ice-cream") to expectations of future ones (e.g.: "if I eat ice-cream now, I will feel this much pleasure"). Present choice (a "want-experience") is just the pursuance of the greatest net pleasure, given one's expectations of the consequences of one's choices (the "like-experiences" we can produce through activity).

To sum up: pleasure is the *sole* object of choice because people want to do what they like, which is what they find more pleasurable[9]. Find out what people like (i.e., they find pleasurable) and you can get to what they want (i.e., what motivates them to act); and, with some luck, given their resources and beliefs, you may get to what they are planning to do.

But how do we compare different pleasures and get to overt conduct then? That is, how do we compare, for example, the pleasure of eating ice-cream, with spending

[9] This is too quick and cursory. Rawls 2000 convincingly shows that Hume didn't think pleasure to be the *sole* object of choice and thus concluded that "Hume's view is neither hedonistic nor egoistic". So, the canonical view is wrong, according to Rawls, which would mean that the real Hume was a lot closer to my manipulation (or would have accepted it more easily) than I argue for. See also Garrett 1997; Castignone 1964; Kemp Smith 1941; Della Volpe 1933; McGilvary 1923.

time with one's partner, or the pride associated with presenting one's research at an important conference, so that we can predict the combination of each activity on the budget line, which is the most pleasurable, and thus will be chosen? Hume gives us (at least) four principles for combining passions together, and estimating their relative strengths[10]:

(i) *The principle of the predominant passion:* this is the tendency of activity to combine different passions together, in which case the weaker passion typically converts into the stronger passion, adding force to it. I may play basketball because I like physical exertion, and because I hate that colleague who plays too and want to beat them to the ground. In which case, Hume would probably predict that, whenever I play with that colleague, my inclination toward the activity increases, as physical exertion combines with hate, with the former converting into latter and the latter ending up dominating and controlling the activity itself.

(ii) *The principle of the greater influence of more particular and determinate ideas on the imagination.* The idea here is that pleasures, which we have more acquaintance with, and more specific and detailed ideas about, generally win against vague and abstract ideas of pleasure and advantage. So, if they offer you to spend the evening with your partner; or to have dinner with some unnamed luminaries in your field, you may choose the former over the latter even when you care more about your career than your partner, because you are able to form a more detailed and specific idea of the pleasure resulting from the former than from the latter. Or, perhaps more generally, people generally prefer public policies when they can put a name and a story to its beneficiaries, over public policies, which merely appeal to abstract and vague (but perhaps bigger) general advantages.

(iii) *The principle of time-discounting.* The idea here is that relatively closer (in time) pleasures count more than relatively more distant ones, so that you may choose to go out for an ice-cream now, even when finishing up your paper instead will give you more income in the future and thus possibly more ice-cream in the future too, because the bigger pleasure provided by the bigger quantity of ice-cream in the distant future is discounted relatively more than the smaller pleasure of the relatively smaller quantity of ice-cream this evening.

(iv) *Finally, the principle of custom or habituation.* Hume says that habits have two main effects on our passions. First, prior activity facilitates present repetition, and second, it increases one's inclinations toward it. So, when playing basketball, as opposed to, say, playing cards, becomes habitual, one is relatively more likely to choose the former over the latter because having played ball in the past has *both* increased one's abilities to perform the activity (and thus decreased its costs) and one's inclination toward it. Now, these two effects would seem to be

[10] See Hume 1978: II, iii, 4 – 9.

unrelated, the first one having to do, in contemporary parlance, with one's skills; and the second one having to do with one's tastes, or preferences. But Hume seems to think that they are intimately connected, nonetheless. Here is one way of thinking about this relation. Prior activity facilitates present repetition, because it allows us to develop skills in performing the activity in question (first part of the habituation principle). But more, or better skills may mean better mastery of the activity, and better mastery of activity, Hume could be understood as saying, add gratification to performance (second part of the habituation principle, first interpretation). Another, and opposite way, to connect the two effects could be this. Prior activity may facilitate present repetition when it harms the skills, which are involved in *other* activities. The agent is then relatively more inclined to perform the activity because the accumulated effects of prior activity have *added* frustration to other activities (second interpretation of the second part of the habituation principle).

Now, this brief catalogue of concepts is not nearly enough for a decent look at the intricacies of Hume's explanation of the emergence of human sociality, government, and the law generally. To get even a quick glance at such explanation, we would need to add a lot more, which would take us far off the trail. But what we said so far already allows us to see rather clearly why a "Habits and threats" conception of obligation appears too simple and intuitive to be refused off the bat, and where it gets into serious trouble with Hart's arguments in chapter 4 of the CL.

Notice, first, that if pleasure is the sole object of choice, then threats and sanctions do indeed immediately become most visible candidates for explanations of rule-governed behavior. Why would anyone sacrifice their pleasure to comply with a social rule? Because they want to avoid a larger reduction of their pleasure when they get caught and punished.

Does this mean that we should expect social agents to constantly adjust their behavior to shifts in the probability of getting caught? No, if we can expect them to have become habituated to threats and then sanctions in case of non-compliance. Repeated exposure to either one may make them more salient (Hume's principle of the more particular over the more general), having witnessed to actual punishments may make fear of getting caught dominate the rewards of illegal activity (this is Hume's principle of predominance), or discount future pain relatively less (this is Hume's third principle of time discounting), or repeated performance may make compliance less costly (Hume's habituation principle), etc…making it false, generally, to say that the existence of a social rule *only* depends on the probability of incurring in a sanction in case of non-compliance.

And again, this is where Hart's arguments in chapter 4 come in: this is all well and fine, but mere exposure and habituation to threats and punishments doesn't provide a standard for social criticism, it reduces one's attachments to social rules to

either strategic behavior or thoughtless routines or mere compulsions and doesn't account for the normativity of social rules. Habits are not social rules, because they cannot explain their internal aspect.

And we can see now where Hume's big problem lies (in this characterization). By taking pleasure as the sole object of choice, he is forced to reduce "want-experiences" to mere expectations of "like-experiences". One further consequence of this is that he cannot account for situations in which the agent wants to do (or refrain from doing) something, and yet *doesn't* expect to experience any pleasure as a reward, because they are applying a standard, which they are attached to, by more than a mere compulsion, a thoughtless routine, or some long-term future personal gain.

When choice collapses "wants" and "likes" through "pleasure", we are at odds to explain how any activity at all can be *voluntary* and *not* concerned with one's own pleasure as well. And we are thus at odds to explain how choice governed by social rules can be more than merely strategic behavior, thoughtless routine, or a form of compulsion[11].

4. Tweaking Hume's Conception of Human Conduct

Let's see now whether when we introduce the two tweaks, we can get any further or say something helpful.

Let's begin by the second tweak: Habits are bundles of impulses, expectations, abilities, and environmental variables (including one's social capital[12]), which reduce the relative marginal cost of cognitive awareness and motivational control over one activity or other[13]. Formally, they define intertemporal complementarities between activities in the past, present, and future[14]. Furthermore, they define marginal rates of substitution between *different* activities in the present, and marginal rates

[11] I owe this point to discussions with Pablo Navarro and Diego Dei Vecchi.

[12] See Coleman (1988) for the classic definition of social capital: a measure of the propensities (the habits) of relevant others toward oneself.

[13] For most recent and comprehensive accounts on "human habits", see Testa, Caruana 2020 (building on pragmatist conceptions of human habits), Hutto, Robertson 2020. For a classic historical account of the relevance of the concept of "habit" in sociology, see Camic 1986; for the difficulties of philosophical theories of action to account for habits, see, for example, Douskos 2017a, 2017b, and 2017c. The definition bears affinities, also, to one long-standing project lead by the late Bruno Celano and Marco Brigaglia, to provide an account of social norms informed by contemporary psychology: see Brigaglia, Celano 2018 and Brigaglia 2016. My Humean manipulation closely aligns with (or, rather, is perfectly representable by the algebra of) Gary Becker's models of habit-formation; see Becker 1996: chapters 1 and 4 especially.

[14] By intertemporal complementarities between activities, I mean this: an effect of past activities (and experiences more generally) on present and future inclinations, through the accumulation or depletion of skills.

of substitution between the *same* activity in the past, present, and future. Whenever Hume writes "original passion", read "impulse". And whenever Hume writes "secondary indirect passion", read this definition of "habits".

Notice that this definition of habits is more capacious than Hume's principle of habituation. These habits include, to be sure, the effects of past activity on present abilities. Repetition of the activity in the past may make present activity easier by relatively increasing the agent's skills (this is the first part of Hume's habituation principle). Also, habits include time preferences too (this is Hume's second principle) since they include marginal rates of substitution between activity in the past, present, and future. But past activity in general may increase our present ability to predict, and vividly imagine, the future consequences of present choice and possibly altering the latter (this is Hume's particular over general principle). Finally, habits may combine different impulses *together*, thus increasing their strength in controlling overt conduct, in two different ways: 1) one impulse may come to dominate all others, (principle of predominance) perhaps increasing inclination toward the activity which it governs, by increasing frustration with the performance of other activities (second part of the habituation principle, second interpretation); 2) or, different impulses may integrate themselves in the performance of one activity, thus increasing one's mastery over the activity which joins them, and increasing inclination toward it (second part of the habituation principle, first interpretation).

Line up all these characteristics together and you get Hume's double relation between ideas and sensations, which contradistinguish secondary, indirect passions. Habits incorporate, first, static judgments. That is, they incorporate marginal rates of substitution between activities (this is the first relation): how much one is willing to forego one activity in exchange of performing another. But they may incorporate dynamic judgments too. That is, they incorporate judgments on the effects of present choices on *future* static judgments (this is the second relation): How much one is willing to forgo present activity in exchange for its effects on future marginal rates of substitution between activities. To sum up: habits are not just Hume's principle of habituation; they are secondary, indirect passions, which combine impulses (original passions, in Hume's vocabulary) together with expectations and skills. They depend on the accumulated effects of past activity, and incorporate static and dynamic valuations of past, present, and future activity.

So, this second tweak appears to be more terminological than anything else. This new definition of "habits" may help to re-organize Hume's vocabulary, and make it fit within a more contemporary jargon, but little else. There are finer contributions too, however.

For one thing, by switching from "passion" to these "habits" as the basis of human activity, we have a chance at understanding "strategic behavior under threat of punishment", as well as "thoughtless habituation" and "compulsions", not simply as different types of behavior, but as special instances of the much bigger phenome-

non of human habits. But for "habits" to become an intelligent criterion for differentiating between these different forms of behavior, we need to show that we can distinguish the latter through descriptions of interactions between impulses, expectations, abilities, and social capital. I pick up (a small part of) this task in section 5.

Second, and consequently, "habits" become a lot more than mere repetitive behavior. They are acquired pre-dispositions to act, and to perceive and respond to *stimuli*. They are not necessarily thoughtless. They influence the total level of cognitive awareness and motivational control, by reducing their relative marginal costs. They are ways of valuing, which may or may not be flexible, and responsive to changes in the environment, depending on the abilities of the agent, their expectations, and their abilities to adjust both to these changes. Section 4 will spell out more details on these definitions.

Let us now look at the first tweak: Hume's pleasure is not the sole object of choice, a "like – experience", an experience at the terminal point of activity, and which motivates it as a reward; Pleasure is a "want – experience", an experience at the moment of choice, and which pushes it forward. Whenever Hume writes pleasure, read stimulation.

At first sight, this may look a lot like hair-splitting. What difference could there possibly be in understanding pleasure as a "want-experience", instead of as a "like-experience" *too*? In many ways, none: a large chunk of predictive statements on human behavior which understand "pleasure" *both* as a "like-experience" (and "pleasure-talk" as "like-talk"), *as well as* a "want-experience" can be translated into statements which understand "pleasure" *only* as a "want-experience" and "pleasure-talk" as "stimulation-talk" with no change in their truth-conditions. Which is good news, because it means that empirical evidence supporting a statement about human behavior expressed in "pleasure-talk" understood as "like-talk" supports, just as well, a statement about human behavior expressed in "pleasure-talk" understood as "stimulation-talk".

But in a few fundamental ways, the difference is theoretically significant. Understand Hume's "pleasure" as a "want-experience" *only*, and "pleasure-talk" as "stimulation-talk", and you can distinguish between "want-experiences" and "like-experiences", *without* assuming that they necessarily tend to converge in human conduct.

Do people always act on what's more "pleasurable" to do? Can we still infer conduct from "competing wants over scarce resources"? Yes, to both, but what we mean by this is simply that people cannot fail to respond to what stimulates them. It doesn't mean that they are always and invariably stimulated by what they like, nor does it mean that what they like is invariably what's more pleasurable *once achieved,* nor that they are invariably aware of what they want, or what they like, nor that they reflectively endorse whatever it is that they either want or like. Put more generally, we do *not* have to assume that people invariably act on a hedonistic interpretation of what being "autonomous", "intelligent", "conscientious", "rational" or "reason-

able" means, but rather we understand these adjectives to refer to different *types* of human habits – being "autonomous", "intelligent", "conscientious", "rational" or "reasonable" respectively mean to have habits of a kind which makes one sensitive to particular classes of *stimuli*.

But if we can distinguish between "want-experiences" and "like-experiences", and thus remove Hume's first big obstacle, do we get a new shot at Hart's three challenges? Or, more concretely: do we get a new shot at using habits to explain social rules? This is the task for section 5.

5. A Fresh Look at Hart's Three Challenges

Let's take stock. So far, I have reviewed (some of) Hart's arguments against a "Habits and threats" theory of social obligation. I began by integrating his arguments in chapter 4 of CL, with his arguments in chapter 7 against "predictive" (and verificationist) theories of social and legal obligation. Habits complement these "predictive" (and verificationist) theories and help them overcome some of the most stubborn difficulties they face (first part of the efficacy challenge). But "Habits and threats" alone cannot account for social obligations either because habits are not social rules. *Inter alia*, social rules are different than habits because they are more than mere repetitive and thoughtless behavior, and they include standards for social criticism (second part of the efficacy challenge); they are also more than mere compulsions (compulsion challenge), because they have normative force (normativity challenge).

Next, this essay looked at the Humean origins of this "Habits and threats" theory of social obligation. It finds that a (candidly manipulative, and somewhat speculative) textual interpretation of the *Treatise* supports a conception of habits and human motivation, which looks promising for explanations of obligations. These two interpretative tweaks to Hume's framework help one to re-organize Hume's key theoretical constructs in a perspicuous way and lend a hand with one of its most daunting problems (i.e., distinguishing between "wants" and "likes").

It is now time to test this emerging conception of social obligation against Hart's challenges. Are the two tweaks well equipped to account for social rules?

Again, the evidence is speculative (but hopefully far less manipulative). To anticipate: the two tweaks perform well on the "efficacy" challenge and the "compulsion" challenge. They only perform decently on the "normativity" challenge; but there will be reason to cheer to that too. Let's begin with a few definitions and see where they lead us.

5.1. Definitions: Habits and Social Rules

Define a rule as the name of a habit; that is, as the name of a secondary, indirect passion, which combine impulses together with expectations, skills, and environmental variables (including one's social capital). These habits depend on the accumulated effects of past activity and incorporate static and dynamic (conscious or deliberate) valuations of past, present and future activity.

A *social rule* (as opposed to an *idiosyncratic* one) is the name of a habit, which includes expectations about the effects of one's choices over one's social capital. A person who possesses such a habit incorporates marginal rates of substitution between different activities, and marginal rates of substitution between past, present, and future activities. Such person, if the habit they possess represents a social rule, is responsive to expectations about the effects of their choices and activities on the choices and habits of other people, and thus their propensities toward them.

A rule *exists* whenever we can give a name to a habit, either real or imaginary, and use its valuations as standards of behavior. An *effective* rule is one whose existence reduces the relative marginal cost of cognitive awareness and motivational control over one activity or other. Or a rule is effective for one person whenever that person possesses the relevant habit. Rules in general, and social rules particularly, can be *shared*, or not, within a social group, depending on whether its members share, or not, the relevant habit.

There are two general ways of making a rule effective – that is, there are two ways in which the existence of a rule generally may reduce the relative marginal cost of cognitive awareness and motivational control over one activity or other. A rule may reduce the relative marginal cost of cognitive awareness and motivational control over one activity by increasing the marginal cost of cognitive awareness and motivational control of all other activities. Call this a "compulsion": after a while, the agent must perform the activity again, because the cognitive and motivational marginal cost of doing anything else is just too dear. This is Hume's principle of the predominant passion, and the second interpretation of the second part of the habituation principle[15].

Or a rule may decrease the relative marginal cost of cognitive awareness and motivational control of one activity *directly*, without increasing the marginal costs of cognitive awareness and motivational control of all other activities. Call this a

[15] Note that this simple strategy allows to account for the opacity of (linguistic formulations of) rules, and pushes a "habits and threats" away from a simple command theory of law: if rules are names of different configurations of human habits which share an effect on people's inclination to perform an activity, but differ with respect to how this effect is achieved, and thus differ with respect to how the agent shall respond to future modifications of the environment, than neither shifts in the probability of incurring in a sanction, nor the mere "acceptance" of the deontic qualification of an activity are enough to predict the agent's future behavior.

"commitment"[16]: the activity is performed because it comes out easy, naturally, and automatically, and the agent can exert (cognitive and motivational) control on its development with smaller marginal costs. This is the first interpretation of the second part of the habituation principle[17].

5.2. On How to Meet Hart's Three Challenges

Enough with definitional posturing though: does this conception of habits help in any concrete way with Hart's three challenges? Let's begin with the "efficacy challenge". The first part of this challenge demands that we distinguish between the existence of a social rule and its judicial efficacy: a rule may exist even when we cannot plausibly predict a sanction in case of violation of the rule itself. We have seen how the received interpretation of a "Habits and threats" conception provides (at least the beginning of) an answer to this challenge. Rule-compliance can be just as much about habituation and thoughtless repetition as about smart adjustments to expectations. But now we can add something more: if a social rule is a name of a habit, then it is the name of one possible *cause* of actual behavior. As we have said, habits causally influence overt conduct by reducing the marginal cost of cognitive awareness and motivational control. And causes cannot be *analytically correlated* with their effects, making any attempt at reducing social rules to overt behavior a vacuous goal.

But habits causally influence overt behavior because they incorporate human valuations too. So, a social rule is a name of a possible cause of overt behavior, and, as such, of a way of incorporating valuations within one's acquired pre-dispositions to act. Furthermore, following a social rule implies the development of habits which include expectations about the consequences of one's behavior on other people's propensities toward oneself – which means that the agent now sees their behavior as part of a more general one and use the latter as a standard for their own behavior as well. These remarks deal with the second part of the efficacy problem: the existence of a social rule is the existence of a name of incorporating valuations, within one's pre-dispositions to act, and which include expectations about the effects of one's

[16] This definition closely aligns, it seems to me, to Garrett's interpretation of commitments in Hume (and to the role they play in facing Hume's skeptical challenges in a humanly responsible way). See Garrett 1997: 237

[17] Note that the distinction between compulsions and commitments provide an explanation of why Hart took it as a matter of course that habits are mere behavioral regularities, and why he was most probably wrong in his diagnosis – repetitions might very well be key in the formation of both compulsions and commitments, depending on their effects on the non-depleted skills of the agent. But whereas compulsions deplete the agent's skills in performing other activities and thus push them to repeat behavioral routines, commitments do not, making the (external) observation of behavioral regularities and routines more difficult (and certainly not necessary for the existence of the relevant habit).

choices on the pre-dispositions (the habits, that is) of relevant others (what I called, one's social capital). A social rule is thus more than merely convergent behavior and thoughtless repetitions; partly because it is a name of *one* possible cause of such convergence, and partly because the efficacy of this cause depends on the acquisition and efficacy of a standard for social criticism.

The second challenge (the compulsion challenge) demands that we make room for more than compulsions in explanations of the internalization of social rules. This is precisely what the distinction between compulsions and commitments does. Compulsions reduce the relative marginal cost of cognitive awareness and motivational control of performing one activity over another, by increasing the marginal cost of the second one. Commitments do the opposite: they reduce the relative marginal cost of the first one, by directly reducing it.

But what about the normativity challenge? I said I only have a decent answer so I should begin slowly, and in the negative. The first contribution is that the two tweaks prevent hedonist utilitarians from using Hume to make a bad political argument. The fact that "wants" are understood to converge with "likes", while the latter are reduced to "pleasure", might appear like a vindication of the principle of (hedonic) utility, as an interpretation of justice. Why should public policy aim at maximizing pleasure? Well, because that's what everyone wants *and* likes anyways!

At least since the beginning of the 20th century, however, many philosophers have been persuaded that powerful arguments by Moore have effectively demolished Mill's (much more sophisticated) version of this argument[18], connecting justice to utility, by connecting "wants" to "likes" (It is, I think, very instructive to notice however that Mill had plenty to say on cases in which "wants" do *not* appear to converge on "likes". And the reader, I suspect, won't be surprised by his examples either: compulsions, habits, and virtuous action) [19].

Bracket Moore's anti-naturalism against Mill however, and what's key is that the first tweak effectively blocks this utilitarian strategy right from the start. By severing the connection between "wants" and "likes", the first tweak breaks open this simple way of connecting justice and utility too. So, the first contribution of this framework is that we are not authorized, without further and penetrating arguments, to infer any advice on reasonable choice from the mere fact that we are observing that utility/stimulation is going up, or down.

The second contribution strikes a more positive note. To see it, think of this very plausible objection to what has been said so far. The characterization of social norms as names of habits of a certain type fails to account for one common intuition. Namely, the intuition that compliance with a social obligation often requires a sacrifice – something one wouldn't do, were it not for the obligation itself. But how

[18] See, for example, Taylor, 1993 for an illuminating account of this classic debate.
[19] See Mill 1861.

can habits represent this "phenomenology of sacrifice" if they reduce the relative marginal cost of cognitive awareness and motivational control, and thus make activity easier, more natural, and decidedly less difficult (relative to its alternatives)?

There are three strategies available. One is to say, most generally, that severing "wants" from "likes" implies that easier action doesn't mean more net pleasure as a result. That is, nothing in this formal representation of habits prevents one from believing that people can be stimulated to sacrifice themselves, for the sake of someone (or something) else, and with no hope of any personal return. Reduction of the relative marginal cost of cognitive awareness and motivational control simply means that the activity stimulates the agent relatively more.

The second strategy is to interpret this "phenomenology of sacrifice" within a very specific moment in the deliberative life of human agents: situations in which an agent perceives uncertainty in present stimulation, and must thus act now, with a view of resolving the uncertainty and possibly changing their (or other people's) future habits too. Notice that this is still a prudential view of human agency. One decides what to do now, by considering the likely future consequences of their choice. But this is a very special kind of prudence however: first, because relevant consequences include the effects of present choice on one's *future* habits as well; second, because the causal imputation of effects shouldn't stop at actions, but rather recede to *choices* as well. Third, and consequently, because the agent chooses a proximate goal (finding out the consequences of present choice), with a remoter goal in view (using such discovery to reform one's habits). The result is that action directed at changing one's own (or other people's) habits may be able to represent part of this "phenomenology of sacrifice", as a peculiar form of *investment* in one's own, or other peoples', habits: the agent accepts costs now, to reform habits (and thus choices and activity) in the *future*.

The third strategy is to say that one crucial component of the "normativity challenge" is a demand to account for how one's acceptance of a social rule may involve a critical reflective attitude toward certain patterns of behavior, and that is precisely what the distinction between "commitments" and "compulsions" does. Recall that, according to my definition, habits reduce the (relative) marginal cost of *cognitive awareness and motivational effort* of performing one activity. So, it is not *just* that the activity is altogether cheaper. Or: the activity is cheaper because it is easier to exercise one's thinking over its performance and control one's motivation to do it. And it is the specific way in which this reduction is achieved which differentiates, as we have seen, "commitments" and "compulsions". Whereas the latter make the use of "thought" to perform the activity relatively cheaper, because the agent cannot think of, and thus cannot motivate themselves to do, anything else; the former make "thought" and "motivation" easier, because the agent has developed better skills and mastery over the activity itself. So, internalization of a social rule through a commitment can account for a "critical reflective attitude toward certain patterns

of behavior". These patterns of behavior are acted upon because they are "easier" and "cheaper", sure. But this means that the agent can thoughtfully monitor, and control the development of activity, be more sensible and responsive to a wider and finer set of consequences, both intended and unintended.

6. Conclusion

Was Hart wrong then in attacking a "Habits and threats" conception of legal obligation? It depends.

If you read Hart's arguments in chapter 4 as a refutation of an entire (Humean) tradition of interpreting social obligations, yes; because there are visible and almost ready at use conceptual and theoretical resources within Hart's (and Austin's) own tradition, which allow for a finer definition of habits. I proposed two tweaks to a canonical interpretation of Hume's conception of human choice and conduct: 1) When Hume writes "pleasure", read "stimulation". 2) When Hume writes "secondary indirect passions", read acquired pre-dispositions, which incorporate both static and dynamic valuations, and which reduce the relative marginal cost of cognitive awareness and motivational control over one activity or other – read habits, that is. And I have shown how these two tweaks have something helpful to say on all of Hart's three challenges.

But does this amount to a refutation of Hart's arguments against a "Habits and threats" theory of legal obligations then? No. For one thing, this essay doesn't propose a complete theory of social obligations and much less a full theory of legal orders and institutions (i.e.: a set of propositions on the origins, development and transformations of social and legal rules, which can withstand empirical testing). This essay is just the beginning of an analysis of a few important building blocks of an empiricist conception of social obligations in general, and of law and legal obligation. And nothing more than that.

Furthermore, nothing here is meant to refute Hart's claim that social rules, legal ones included, are more than mere compulsions and repetitive behavior. This essay merely adds that we can understand habits to be more than these things too. When we come to see that, this essay further infers, we have a rich (albeit incomplete) model for studying social obligations generally, and perhaps legal ones specifically, in the wild. More implicit, but we also have a message of hope for the hard-nosed realist: you do not have to reduce all human attachments to either pleasure-seeking strategic behavior, compulsions, or thoughtless habituation, even if you want an empiricist conception of social obligation (save when either one of the former reductions provide a better explanation of social behavior)[20].

[20] To put it bluntly: by accepting these two tweaks to Hume's conception of human motivation

Finally, consider the normativity challenge once again. As we have seen, Hart argues that a theory of legal obligation must explain law's normativity – An account of when, that is, law's authority binds its subjects. I said I merely have a decent reply here and shall thus be brief in recapitulating why: the major advantage of this conception of habits and human motivation is that it seems compatible with a host of different normative approaches. Somewhat surprisingly, and its origins notwithstanding, the two tweaks to the Humean conception of human choice and conduct do not easily collapse into any form of (hedonist) utilitarianism. Pleasure is still all-important in this approach, to be sure, but not as the sole object of choice – it is stimulation.

And that's good news too: it means that the theoretical vocabulary proposed here can be shared by different and perhaps otherwise incompatible moral and political traditions. If you think that legal and political theory should help people reduce the risk of talking past each other in legal and political discussions, then perhaps you agree that a shared theoretical vocabulary is an asset, albeit a clearly incomplete one.

References

Austin, J. (1831). *The Province of Jurisprudence Determined*, London, John Murray.

Becker, G. (1996). *Accounting for Tastes*, Cambridge (Mass.), Harvard University Press.

Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*, Cambridge (UK), Cambridge University Press.

Brigaglia, M., Celano, B. (2018). *Reasons, Rules and Exceptions*, «Analisi e Diritto», 131-144.

Brigaglia, M. (2016). *Rules and Norms. Two Kinds of Normative Behaviour*, «Revus», 30, 33-57.

Castignone, S. (1964). *Giustizia e bene comune in David Hume*, Milano, Giuffrè.

Celano, B. (2016). *Pre-conventions. A View of the Background*, «Revus», 30, 9-32.

Camic, C. (1986). *The Matter of Habit*, «The American Journal of Sociology» 91, 5, 1039-1087.

Coleman, J.S. (1988). *Social Capital in the Creation of Human Capital*, «The American Journal of Sociology», 94(S), 95-120.

Coleman, J L. (1982). *Negative and Positive Positivism*, «Journal of Legal Studies», 11, 139-164.

and practical agency, one doesn't have to turn "hermeneutic" to recognize the importance of Hart's three challenges or, alternatively, become a "hedonist utilitarian" to find insight in Hume's empiricist framework.

Coleman, J.L. (2001). *The Practice of Principle*, Oxford, Oxford University Press.
Dewey, J. (1894). *Austin's Theory of Sovereignty*, «Political Science Quarterly», 9, 31-52.
Dewey, J. (1922). *Human Nature and Conduct: an Introduction to Social Psychology*, New York, Henry Holt and Company.
Dickson, J. (2007). *Is the Rule of Recognition Really a Conventional Rule?*, «Oxford Journal of Legal Studies» 27, 373-402.
Douskos, C. (2017a). *Pollard on Habits of Action*, «International Journal of Philosophical Studies» 25, 4, 504-524.
Douskos, C. (2017b). *Habit and Intention*, «Philosophia» 45, 3, 1129-1148.
Douskos, C. (2017c). *The Spontaneousness of Skill and the Impulsivity of Habit*, «Synthese» 196, 10, 4305-4328.
Dworkin, R. (1977). *Taking Rights Seriously*, Cambridge (Mass.), Harvard University Press.
Garrett, D. (1997). *Cognition and Commitment in Hume's Philosoph*, Oxford, Oxford University Press.
Gilbert, M. (1990). *Rationality, Coordination and Convention*, «Synthese», 84, 1-21.
Green, L. (1999). *Positivism, and Conventionalism*, «Canadian Journal of Law and Jurisprudence», 35-52.
Guala, F., Hindricks, F. (2015). *A Unified Social Ontology*, «The Philosophical Quarterly», 65, 177-201.
Guala, F. (2016). *Understanding Institutions: The Science and Philosophy of Living Together*, Princeton, Princeton University Press.
Hart, H.L.A. (1961). *The Concept of Law. 3d Edition* (2012), Oxford University Press.
Hume, D. (1978). *Treatise of Human Nature*, P.H. Nidditch (ed.), Oxford, Oxford University Press.
Hutto, D. D., Robertson, I. (2020). *Clarifying the Character of Habits*, in Caruana, F., Testa, I. (eds.). *Habits: Pragmatist Approaches from Cognitive Science, Neuroscience, and Social theory*, Cambridge, Cambridge University Press, 204-222.
Kemp Smith, N. (1941). *The Philosophy of David Hume: A Critical Study of its Origins and Central Doctrines*, London, Palgrave MacMillan.
Krygier, M. (1982). *The Concept of Law and Social Theory*, «Oxford Journal of Legal Studies», 2, 155-180.
Leiter, B. (2007). *Naturalizing Jurisprudence. Essays on American Legal Realism and Naturalism in Legal Philosophy*, Oxford, Oxford University Press.
MacCormick, N. (2007). *Institutions of law: an Essay in legal theory*, New York, Oxford University Press.

McGilvary B. (1923). *Altruism in Hume's Treatise*, «The Philosophical Review», 12, 272-298.

Mill, J.S. (1861). *Utilitarianism, in Collected Works of John Stuart Mill* X, J. Robson (ed.), Toronto, Toronto University Press.

Moles, R.N. (1987). *Definition and Rule in Legal Theory: a reassessment of H.L.A. Hart and the positivist tradition,* Oxford, Basil Blackwell.

Postema, G. (1982). *Coordination and Convention at the Foundations of Law*, «The Journal of Legal Studies», 165-203.

Rawls, J. (2000). *Lectures on the History of Moral Philosophy*, Cambridge (Mass.), Harvard University Press.

Taylor, C. (1993). *Explanation and Practical Reason*, in M. Nussbaum and A. Sen (eds.), *The Quality of Life*, Oxford, Oxford University Press.

Testa, I., Caruana, F. (2020). *The pragmatist reappraisal of habit in contemporary cognitive science, neuroscience, and social theory*, in F. Caruana, Testa, I. (eds.), *Habits: pragmatist approaches from cognitive science, neuroscience, and social theory*, Cambridge, Cambridge University Press, 1-40.