It's impossible to will to be punished? Exploring consensual way out of the Kantian dilema

Parmigiani M.

SECRETARÍA DE INVESTIGACIÓN

UNIVERSIDAD SIGLO 21

Revista de
Estudios
Kantianos

Revista de
Estudios
Kantianos

Revista de Estudios Kantianos

Publicación internacional de la Sociedad de Estudios Kantianos en Lengua Española
Internationale Zeitschrift der Gesellschaft für Kant-Studien in Spanischer Sprache
International Journal of the Society of Kantian Studies in the Spanish Language

Diseño, editor de estilo y maqueta
Josefa Ros Velasco, Universidad Complutense de Madrid

Entidades colaboradoras
Sociedad de Estudios Kantianos en Lengua Española (SEKLE)
Departament de Filosofia de la Universitat de València

Índice

Artículos

Recensiones

Informes

Entrevistas

Convocatorias y normas para autores

Artículos

Is it "impossible *to will* to be punished"? Exploring a consensual way out of the kantian dilema

MATÍAS PARMIGIANI[1]

Abstract
In the *Metaphysics of Morals* Kant wrote that "it is impossible to will to be punished". The main goal of the present paper is to challenge this idea. In contemporary literature, a similar challenge was attempted by assigning a pivotal role to the notion of 'consent'. Therefore, focused on these antecedents, what I will try to do in this paper is to determine whether the notion of consent is capable of playing any role whatsoever in a justificatory theory of punishment.

Key words: Kant, punishment, consent, Nino, Finkelstein

¿Es "imposible querer ser castigado"? Explorando una salida consensual del dilema kantiano

Resumen
En la *Metafísica de las Costumbres*, Kant escribió que "es imposible querer ser castigado". El objetivo principal del presente trabajo consiste precisamente en desafiar esta idea. En la literatura contemporánea, un desafío similar fue asumido de la mano de la noción de 'consentimiento'. Pues bien, tomando como base estos antecedentes, lo que intentaré hacer aquí es determinar si la noción de consentimiento está capacitada para desempeñar algún rol en una teoría justificatoria del castigo.

Palabras clave: Kant, castigo, consentimiento, Nino, Finkelstein

I. Introduction: Facing Kant's Challenge

The idea that someone might will to be punished probably strikes us today as strange, nonsensical and counterintuitive as it must have struck people over 5000 years ago. With few exceptions, crimes tend to be anything but public, precisely because it is in the interest of the criminal not to be discovered. Crimes and secrecy go hand in hand. Concealment is what you reasonably need if you want to preserve untouched the benefits of your own offense. And since punishment seems to cancel these benefits, anonymity is what you will be looking for to avoid being caught and punished, as the Book of Job wisely put it in those early times:

> There are those who rebel against the light, who do not know its ways or stay in its paths. When daylight is gone, the murder rises up, kills the poor and needy, and in the night steals forth like a thief. The eye of the adulterer watches for dusk; he thinks, 'No one will see me', and he keeps his face concealed. In the dark, thieves break into houses, but by day they shut themselves in; they want nothing

[1] Centro de Investigaciones Jurídicas y Sociales (Universidad Nacional de Córdoba) – CONICET.

Revista de Estudios Kantianos
Vol. 2, Núm. 1 (2017): 60-88
ISSN-e: 2445-0669
DOI 10.7203/REK.2.1.8813

> to do with the light. For all of them, midnight is their morning; they make friends with the terror of darkness (24).

Such a generalized idea happens to have kept its philosophical intuitive appeal as it travelled across time. In a famous passage from *The Metaphysics of Morals*, trying to take some distance from MarcheseBeccaria's position that those punishments (e.g. death penalty) that cannot be consented to in an original civil contract are wrongful, Kant vigorously endorsed it. He wrote: "No one suffers punishment because he has willed *it* but because he has willed *a punishable* action; for it is not punishment if what is done to someone is what he wills, and it is impossible *to will* to be punished" (MS 6:335).

Back to our time, the idea seems indisputable in the philosophical literature on the justification of punishment, with two *apparent* exceptions. As we will see in more detail in Part§3, there are at least two philosophical approaches that decide to rest on the notion of consent to justify punishment. Nevertheless, given that none of them really have the intention of equating 'consent' with 'the expression of an individual's real will', or anything like it, in the end they both would prove to gowith the flow. In spite of its compelling character, the main goal of the present paper is to challenge the idea captured in Kant's claim. Now, what does it precisely mean to assume this challenge? What are its implications? To give a straight approximate answer that may be used as a guide from now on, it just means to try to determine whether the notion of consent is capable of playing any role whatsoever in a justificatory theory of punishment.

As abstract as it is, the question doesn't seem to contain anything of value. Different concepts of consent will be able to play different roles in different contexts, depending on what we purport to justify and how we pretend to do it. By the same token, there are certain contexts in which the concept of consent, no matter how it is defined, will not be able to play any justifying role. Thus, for example, if the normative order against whose background consent was given is morally bankrupt, "the moral magic of consent – as Kleinig adverts – will not work" (2010: 21).[2] Therefore, in order to reduce the scope of the question, a safer and more appealing place needs to be sighted. Personally, I think this place is provided by the next opposition: on the one hand, what we may call a weak notion of consent, such as the one that sometimes happens to be at work in contract law as well as in tort law; and, on the other, a stronger notion of consent, understood as the expression of an individual's real will.

The paper is divided into three parts and a conclusion. In the next two parts (§2 and §3) I will explore the main attempts that *can be* and *have been* made to allow the concept of consent play a role in a justificatory strategy of punishment. Though we might be able to obtain from their analysis important conceptual tools to move forward, for different reasons none of them will prove to be successful. In part §4 I present the stronger notion of consent and argue as if it were the only one in shape of providing the kind of unique and stringent justification that the infliction of punishment usually demands. After looking for some evidence in fiction stories as well as in real-life cases, I hope to show that an agent that wills to be punished is perfectly conceivable. Nonetheless, this circumstantial evidence under no concept shall lead us to believe that unless someone consents to her punishment, punishment would not be justifiable in her eyes, for this would make us face an unsolvable dilemma. Therefore, in order to bite the bullet without putting us in an untenable position, I would have to come up with anew conceptual strategy. Parts §4 and §5 (*Conclusion*) are meant to shed some light on it.

[2] In a Kantian vein, see Korsgaard 1986.

II. Why not the Esay Way Out?

II.1. Rawls's "Two Concepts of Rules": Delineating the Esay Way Out

At least in one sense, to develop an explanatory strategy that successfully defies Kant's challenge is not a difficult task at all. It suffices if we manage to define 'punishment' in such a way that it leaves it virtually devoid of its most typicallystringent features. For the sake of clarity, I propose to consider such a moveas the easy way out of the dilemma. Unsound as it may look, the strategy has nevertheless left its mark in the philosophical tradition. Fortunately, there is no need to look far back into history, for something very much like it was outlined by John Rawls in "Two Concepts of Rules". Following Hobbes' definitionof punishment given in *Leviathan*, Rawls explains that a person is said to properly suffer punishment whenever the next provisions are satisfied:

(1) He is legally deprived of some of the normal rights of a citizen on the ground that he has violated a rule of law;

(2) The violation *[has]* been established by trial according to the due process of law;

(3) The deprivation is carried out by the recognized legal authorities of the state;

(4) The rule of law clearly specifies both the offense and the attached penalty;

(5) The courts construe statutes strictly;

(6) And the statute was on the books prior to the time of the offense (1955: 10).

For anyone familiar with the philosophical literature on punishment, the list ofprovisions accounted by Rawls in this passage can hardly be taken as conceptually *sufficient*. In particular, there are two missing conditions that cry out for admission: on one part, as Kant already knew, it must be granted that punishment be proportional to the offense (Kant convincingly argues in favor of this condition when discussing the *iustalionis* law) (*cf.* MS 6:332); on the other part, punishment must express some sort of moral condemnation on the person of the wrongdoer (*cf.* Feinberg 1971: 95-118).

For the moment, I will set aside this second condition (*cf. infra*). As regarding the first one, it is interesting to note that Rawls himself would have recognized its importance. In footnote nº 14 (1955: 12) of that very essay, he refers to the principle of proportionality as a requirement of justice that any system of punishment should abide with, trying to offer a response to David Ross's objection that such a principle would be conceptually incompatible with a purely utilitarian justification. Rawls argues that even if this principle did not havea place within utilitarianism *as a requirement of justice*, utilitarianism would be able to account for it in plainly *prudential terms*: apparently, imposing on someone a penalty much severer than his/her offense would be just as irrational as to use a cannon to kill a mosquito, and utilitarianism does always recommend to accommodate our preferences so they can be directed to the less serious possible evil (*cf.* Rawls 1955:12-13).

In Rawls'ss view, then, plainly prudential considerations seem to be in perfect shape to account for the requirement of justice. Anyway, assuming that such a move were accepted, the problem would still be how to pair off an offense and a penalty in such a way that a rational agent, situated under conditions of perfect enforcement, would not still be encouraged to break the law.As might be seen, the prudential coin chosen by Rawls exhibits a second prudential side. How severe do penalties have to be to deter possible offenders under conditions of perfect enforcement? As we know, Kant had a very reasonable answer to this question, which rested on the law of retribution (*iustalionis*). Nonetheless, it is not that answer that should distract us right now. Then, let us keep focusing on Rawls'ss explanatory scheme, for it is from it that there can be derived what I meant to call the easy way out of the Kantian dilemma.

II.2. On Willing and Prices

In Rawls'ss scheme, punishment works just like any other *behavior control system* that may employ "threats to deter with various pragmatic or moral side-constraints on the execution of threats" (Hill 2000: 193). Rawls is explicit on this point when he says, precisely, that "punishment works like a kind of price system: by altering the prices one has to pay for the performance of actions it supplies a motive for avoiding some actions and doing others" (Rawls 1955: 12).At first glance, the actions that the system of punishment supplies a motive for doing are supposed to be precisely those that imply no penalty whatsoever, which is what a commonsensical reading of this passage would intuitively suggest. However, this is not the whole story. If punishment behaves like any price system and a criminal penalty is nothing but one kind of price among others, then there is nothing in Rawls'ss account to skip the inference that an individual may be willing to pay exactly the very same price fixed for a criminal offense.

Moreover, given the fact that punishment behaves like any price system, it is reasonable to assume that the same analytical tools that we employ to understand the *willing behavior* of any rational agent will be equally available when it comes to dealing with the *willing behavior of a would-be punishable* agent. To see that there is no contradiction in this approach, I propose to consider a seven-step practical syllogism ***(PRS)***, purported to shed some light on what it generally means for an agent to want (or will) something that can imply to pay a price which one may not be willing (or consenting) to pay for its own sake. Thus:

(PRS) S1: *If* agent***X*** wants (wills) ***P***;

S2: It is highly unlikely that ***P*** be accomplished without implying ***Q***;

S3: In those rare circumstances in which ***P*** implies ¬***Q***, ¬***Q*** implies for ***X*** to bare such costs (measured in terms of actions,deprivations, responsibilities, sufferings, etc.),that ***P***turns out to be much less rewarding for her;

S4: It is highly unlikely that ¬***P*** implies ***Q***;

S5: There are compelling reasons to believe that ***Q*** can be afforded by ***X***;

S6: And, of the two next available scenarios ***a*** [***P*** and ***Q***] and ***b*** [¬***P*** and ¬***Q***], ***X*** prefers ***a*** rather than ***b***;

C: *Therefore*, ***X*** wants (wills/consents) ***Q***.

A few remarks are in order regarding how to read ***(PRS)***. Conceived in a purely utilitarian vein, ***P*** will be generally represented as a certain state of affairs (*cf.* Williams 1997: 105). Of course, there might be other interpretations available over the table (*cf.* Foot 2002: 95-96; Sen 1979: 464-465). But, even when therewere not, I think that the formula will receive a much more fruitful interpretation if the state of affairs it refers to is understood as the result of my involvement in certain kinds of activities. So, let ***P*** be, for example, the feeling of well-being that subjectively accompanies me as the result of practicing yoga. What I want, then, is not only a certain state of the soul; rather,it is the state of the soul that comes about as a result of doing a certain type of activity, that of practicing yoga. In relation to ***Q***, let it be the price that I have to pay to hire a certain instructor, whom I know from personal experience can provide me like no other with the kind of reward I expect. Being these the main ingredients of the story, we can arrive at the conclusion that it is impossible for me to obtain the rewarding state of the soul I expect from practicing yoga (***P***) without paying what the instructor

ISSN-e: 2445-0669
DOI 10.7203/REK.2.1.8813

charges me as an exchange for his services (***Q***).[3]If I want (will) ***P*** (S1) and ***P*** implies ***Q*** (S2), therefore, assuming that conditions S3…S6 are met, I want (will/consent) ***Q*** (C).

Probably the most important principle that *inspires*a practical syllogism like ***(PRS)*** was raised, again, by Kant.[4] In the *Groundwork for the Metaphysics of Morals* he wrote: "Whoever wills the end also wills (insofar as reason has decisive influence on his actions) the indispensably necessary means to it that are within his power" (GMS 4: 417). Based on this principle, Kant did also emphasize the fruitfulness of the strategy that compels us to interpret our ends as states of affairs that involve us in their production. Thus, he said that "it is one and the same thing to represent something as an effect possible by me in a certain way and to represent myself as acting in this way with respect to it" (GMS 4: 417). What he would have had to say, however, regarding each of the conditions (S2…S6) that must be met for conclusion (C) to be drawn is quite a different matter.

Details shall not distract us here. Once we set them apart, the key element is brought about by premise S5. In order to be willed, a state of affairs and each and every single one of the costs that it involves (deprivations, actions, responsibilities, sufferings, etc.) must be affordable by the agent. In the example given above, what this just means is that I must be absolutely capable of paying the instructor's fare; that is, despite how painful and troublesome the costs of affording it might be, they cannot be as high as to force me to reconsiderthe end's worthiness. To put it in Kantian terms, paying the fare (if that were the only cost implicated here, of course) should be *within my power*.

Now, turning back to the Rawlsian conception of punishment, some important clarifications need to be introduced. On this approach, punishment behaves like any other penalty, except for the six provisions that define its nature (*cf. supra*). At the same time, a penalty would tend to behave like any other price system, were it not for the fact that the application of penalties (especially law-like ones) is generally ruled by a centralized structure very much like the one governing the application of punishment, whereas prices (especially in a free-market economy) respond to different standards. In order to be justifiable applied, however, a penalty does not always have to meet the Rawlsian provisions. Due in part to this sort of deeper liberalized nature that penalties and prices would approximately share, the Argentine Criminal Law Professor, Eugenio R. Zaffaroni, believes it necessary to keepthese concepts cautiously apart from the concept of punishment (*cf.* 2005: 71-72). For my part, and as long as the Rawlsian provisions are clearly set forth in advance, I have no quarrels with such an analogical way of understanding punishment. In fact, the notion of 'price' possesses such an open texture and malleability that it can come as no surprise to realize how it managed to function perfectly well in the widest possible range of contexts. However, once we get seduced by its fertility and assume the challenge to play with it in the criminal law context in particular, some precautions will have to be taken.

For instance, it would look mandatory to say at least a couple of words on the specific *pricy* nature of penalties. In modern economy, as has been extensively explained (*cf.* Barnett 1992: 845; Kotler 1973: 408-409), money is what tend to govern most of our transactions, mainly because its quantitative character, its lack of ambiguity and its one-dimensional nature make it really easy for us to know for sure what to expect from each other. In addition, when a price is fixed for a product (or a service) and the product is in perfect shape, the transaction will be completed as soon as the price is paid. In the case of punishment, however, things are far from being that simple. When someone is sanctioned not merely with an economical fine (or with the loss of a license to run a business, for

[3] Many other examples can similarly do the job. For instance, let it ***P*** be to acquire a certain house in *Park Migdia*. Given that in this park I spent most of the happiest years of my childhood, living in that house would fulfill like no other alternative the dream of my life. Let it ***Q*** be interpreted, on the other hand, as paying its market price, which is 500.000 euros. Conclusion: it is impossible to fulfill the dream of my life without paying the house's market price.

[4] Originally I have chosen the verb 'to govern' instead of 'to inspire' to express the idea. Nonetheless, it is important to emphasize that Kant's principle is, so to speak, of a different scope than ***(PRS)***. For further details on this point, see footnote 12 below.

example), but with a curtailment of his freedom, what this would really imply can sometimes be very hard to tell. But even if that were not the case, what are we supposed to do with the expressive nature of punishment? Are we supposed to interpret it just as another typical transactional device? Regrettably, the expressive feature that is present in most of our forms of punishments does not seem to have a parallel in any other transaction. Moreover, it would be very odd if the blaming function of a specific kind of penalty such as is executed on a particular convicted felon would necessarily have to be interpreted as something that s/he must have accepted beforehand.

II.3. An Analogy between Penalties and Prices

Be it as it may, I truly believe that these important differences are no impediment to keep on strengthening the nuances of the analogy. It must be kept in mind that what we are trying to find out is a clear case of an individual that might be reasonably said to will (or consent) to be punished. Subsequently, once it is admitted that punishment behaves like a price system and prices are fixed on utilitarian grounds, a wide range of credible scenarios is open, in which there is always a chance of finding someone exhibiting a will to be punished. In *MS*, Kant gives many illustrative examples. In one of them, he urges us to consider the case of two men that may have taken part in the 18th Century Scottish rebellion. One of them is a man of honor, whereas the other is a scoundrel. Kant says that if the penalty fixed for taking part in the rebellion had been convict labor, the scoundrel would have had more than one reason to peacefully embrace it (*cf.* 6:334). In a previous paragraph, he criticizes fines as the penalties to be imposed for verbal offenses with the argument that someone wealthy enough "might indeed allow himself to indulge in a verbal insult on some occasion" (MS 6:332). In this example, as will be easily seen, the rich man might indeed will (consent) to be punished, for here punishment seems to performthe same function that ***Q*** performs in our practical syllogism ***(PRS)***, namely –to say the least: punishment looks as unproblematically affordable by him (S5 clause). Unlike Kant, in Rawls'ss scheme cases like these are not at all hard to conceive.

At this stage of the exposition, what we need to do is to explain why such an easy way out of the Kantian dilemma does not look very promising. To do that, I propose to introduce an additional distinction that seems to be concealed in Rawls'ss scheme. Not surprisingly, it is a kind of distinction that Rawls himself succeeded in making during the years of philosophical labor that followed his 1955's essay and crystallized in *A Theory of Justice* (1971). In the shape I would like to introduce it, however, part of the credit goes to H.L.A. Hart. In "Bentham on Legal Rights" (1973), Hart makes a distinction between a concept of a legal right limited to "those cases where the law […] respects the choice of individuals" and a more demanding concept of a legal right, designed to operate as the platform from where to monitor and morally criticize the law. Such a concept of right, Hart says, "is inspired by regard for the needs of individuals for certain fundamental freedoms and protections or benefits," which would not have been paid due care and attention by Bentham's utilitarian approach, exclusively concerned with the problem of how the law was supposed to "maximize aggregate utility" (1973: 200).

Following Hart, Rawls'ss himself and some others (*cf.* Nino 1991), I would like to distinguish between what we may call *non-negotiable fundamental rights*, on the one hand, and *negotiable non-fundamental rights*, on the other. As clear examples of the first kind of rights, there can be mentioned, for instance, all the basic liberties that Rawls derived from his First Principle of Justice, such as the political liberty to vote and run for office, freedom of speech and assembly, liberty of conscience, freedom of personal property and freedom from arbitrary arrest (*cf.* ibid: 53), among others. As clear examples of the second kind of rights, on the other hand, there can be mentionedmany of the most common or even trivial rights that I may have, such as the right to buy a ticket to fly to Hong Kong in business class or the right to have a date this very evening with the person I love. Some of these rightscan surely be derived from the liberties consecrated by the First Principle of Justice. Nonetheless, there are some other rights, like the liberty to "own certain kinds of property (e.g. means

of production)" and the "freedom of contract as understood by the doctrine of laissez-faire" (ibid: 54), which are not derivable in this sense and are not, for the same reason, protected by the First Principle.

Without pretending this classification to be exhaustive, there seems to be a great difference not only in the way these kinds of rights are philosophically grounded, but also in the way they must be balanced to protect individuals against illegitimate hindrances of their freedom. In Rawls's *Theory of Justice* impressive account, whereas the list of primary goods is obtained as the result of a mental experiment that leaves mere considerations of preference satisfaction out of the picture, there is nothing there to prevent that at least an important part of the non-fundamental rights that a person must be recognized as having can be defended on a simple preference-based account of human welfare. After all, as Richard Arneson put it, once a fair background of fundamental human rights is granted, "each individual is responsible for developing a set of final ends and a plan of life to achieve them and for organizing her own life to satisfy this plan" (2000: 236). At this point, utility-based considerations are widely accepted.

Hence, when it comes to realize how to conceive a penalty, the deprivation of rights that it implies by definition (see *supra*, Rawls'ss provision 1) can be accounted for in two ways, as the aforementioned considerations seem to suggest: a) it can be accounted on a model of negotiable non-fundamental rights; or b) it can be accounted on a model of non-negotiable fundamental rights. If we choose the first model, a typical penalty will take the form, for example, of a deprivation of individual bodily movement's freedom, which is one of the most universal forms of punishment. On the contrary, if we decide to take side with the second model, typical penalties may be conceived in the form of fines, restitution, seizure of assets, and the like.

Assuming that a deterrent justification of punishment has some appeal, each alternative modeltaken in isolation will give rise to the next possible conceptions of penalties:

> **A) Penalties that trivially deter**, because the price they fix for an action is established with the help of a negotiable non-fundamental rights model (grounded on a preference-based criterion of human welfare), and this price is highly enough as to discourage a reasonable agent provided with certain rights, preferences and resources.
>
> **B) Penalties that do not trivially deter**, because the price they fix for an action is established with the help of a non-negotiable fundamental rights model (grounded on a non-preference based criterion of human welfare), and this price is highly enough as to discourage a reasonable agent provided with certain rights, independently of her preferences and resources.

Related to **A)**, it must be said that the deterrent function of a penalty grounded on a mere preference-based criterion of human welfare can be thought to be accomplishedmainly with regard to agents whose preferences and resources do not allow them to afford certain prices. Thus, be it a $ 10.000 fine the penalty at stake for insulting someone. In conditions of perfect enforcement, the threat that it involves will mainly work for agents who, in spite of willing to break the law, lack the economic resources that would allow them to afford its cost.For these agents, such a penalty can hardly ever be a welcome consequence. That is why, in a scenario like this, it would be almost impossible to observe non-wealthy law-breaking people. Of course, if an agent possesses not many economic resources but sufficient enough to face the $ 10.000 fine, once she decides to break the law it would be reasonable to conceive her willingness to be punished as a function directly derived from her own preferential background. Notwithstanding her relatively low level of resources, it is likely that she prefers a state of affairs in which she insults someone and pays the fine [**P** and **Q**] rather than one in which she does not commit the offense but keeps the money [**¬P** and **¬Q**]. Meanwhile, for a wealthy law-breaking agent in conditions of perfect enforcement, the willingness to be punished seems indisputable.

Conversely, in conditions of less than perfect enforcement, things are not that clear. In those conditions, the meaning of human acts tends to be highly and, sometimes, unyieldingly ambiguous (*cf.* Barnett 1992: 902). There, what both wealthy and non-wealthy agents may be said to be willing or preferring when they decide to break the law will usually be a purely speculative matter. Regarding non-wealthy agents, considerations of rationality will generally suggest that they are acting with the hope of circumventing the penalty. But what are we supposed to infer from the non-law-abiding conduct of wealthy agents? In conditions of uncertain enforcement, their actions may or may not signify a willingness to be punished. Likewise, what both wealthy and non-wealthy agents may be said to be willing or preferring when they decide to abide by the law will also tend to be a mere speculative matter. Just as there are prudential reasons, there are non-prudential or moral reasons as well. Ergo, it is possible that both wealthy and non-wealthy agentscould be moved from time to time to abide by the law when their preferences and resources push them in the opposite direction.

In contrast to **A)**, the **B)** conception sees universal deterrence as a perfectly accomplishable enterprise, at least in theory. Be it, for instance, a one-year prison term the penalty to be imposed on whoever insults someone. In conditions of perfect enforcement, who could really dare to insult someone and be punished? The immediatereply would read: anyone who prefers a state of affairs in which she insults someone and receives a one-year prison sentence [**P** and **Q**] rather than one in which she abides by the law and keeps her freedom [¬**P** and ¬**Q**]. Yet, multiply that amount by two, by three, or by any number you want, and sooner or later you will find a penalty that no reasonable agent can be taken to assume, at least not on the basis of her *trivial* preferential background. Further in this paper, we will see the importance of this last clarification (*cf. infra*). For the moment, it suffices to note why, according tothis conception of penalties, real personal preferences and resourcescan be accounted as almost practically irrelevant to determine how punishment must accomplish its proper deterrent function.

With regard todetermining the precise amount of a penalty for a given offense, the non-trivially deterrent conception recognizes a distinctively human characteristic that is totally absent in the other conception. In *A Theory of Justice*, Rawls built a philosophical device (e.g. the Original Position) to try to expose how it might be the case that the reasons to recognize a basic set of rights cannot rest upon mere preferential evaluations. Even if that account was far from raising total agreement among philosophers, other compelling arguments have been given to acknowledge such a basic set.[5]On these accounts, the set of rights that we cannot do withoutis precisely the one that defines the autonomy of a person, being its presence the minimum standard requirement that must be granted for a human life to be worth living. Without autonomy, no one would be in the condition neither of pursuing the most ordinary life-plan, nor even to form a preference set order and live in accordance with it.[6]

Hereof, as soon as we are persuaded by the force of any of these arguments (and I truly believe that there are compelling reasons to be thus persuaded, though I cannot go here into further details), there seems to be similar powerful reasons to recognize that, in conditions of total enforcement, a system of punishment organized on a conception of penalties molded by a non-negotiable fundamental rights' model is perfectly capable of accomplishing its deterrent function beyond human beings' contingent preferences. After all, how strong would have to be a preference to

[5] The list of arguments that have been given in the philosophical literature is immensely huge. Some of the most deeply discussed, for instance, appeared in: Raz 1979; Sen 1982; Höffe 1989; Nino 1991; Barry 1995; Wall 1998; Scanlon 1998. But there have been others indeed. Here it suffices to note that the need to recognize a basic set of rights as defining the contours of an autonomous agency has been widely acknowledged, although there reigns no unanimous agreement over which specific rights must be part of this set. For my argument, that contention has no further implications.

[6] In *Not Just Deserts* (1990), this basic set of inalienable rights is conceptualized as constituting not the freedom (or autonomy) but the *dominion* of a person, which is a notion deeply embedded in the republican tradition of political philosophy. I will have the chance to rescue additional elements form Braithwaite and Pettit's republican theory of criminal justice along subsequent sections of this paper.

ISSN-e: 2445-0669
DOI 10.7203/REK.2.1.8813

motivate someone to withdraw nothing less than what enables her not only to form most of her preferences but to live by them?

In conditions of partial enforcement, it is out of discussion how non-trivial penalties accomplish their deterrent function. There too will one observe the presence of wealthy and non-wealthy agents trying to break or abide by the law. And there too will often be their preferences and motives (prudential as well as non-prudential ones) the ones that decide the bestcourse of action. When moral reasons are absent and a preference set order motivates a law-breaking conduct, what happens next is always a complex matter. If the threat of punishment istoo weak, there will be probably too many agents disposed to break the law. If the threat is stronger, there will be fewer. How many law-breakers there could be, however, will always be a varying issue, depending on many intervening factors. Amongst them, there can be mentionedfrom *personal conditions* determining the agents' resources ortheir level of risk aversion, for instance, to *institutional conditions*, such as the effectiveness of the police authorities to catch criminals or the way the procedural criminal system works. All of these factors operating together will contribute to determine the price that a would-be offender is able to pay within a system of punishment in conditions of partial enforcement. That price, indeed, will be unknown most of the time. But the most important corollary of this section has to do with what can be said of an offender's presumed willingness when she commits an offense. And I am sure that the intuition many of us deeply share is that, in a system like this, it is highly unlikely that an offender might be taken to express a will (or consent) to be punished.

II.4. After All, Why not the Easy Way Out?

In Rawls'ss scheme, where punishments are equated with penalties and penalties are equated with prices, there is not the slightest chance of telling what exactly is that distinguishes a *criminal* system of penalties enforcement from other penalties enforcement systems, such as those that are so common in contract law and tort law. The reason for this silence has to do in part with something already seen at the beginning of this section (*cf. supra*, 2.1), when it was suggested that Rawls'ss definition of punishment omits two important features that cry out for recognition: an adequate principle of proportionality and the assignation of an expressive function to penalties.

In relation to the principle of proportionality, what I tried to show is that Rawls'ss utilitarian scheme indirectly accounts for it (*cf. supra*, 2.1). An adequate account, however, would require not only to establish how much is too much but also how much is too low regarding the nature and amount of a penalty, and here is where Rawls'ss account fails. Again, when punishments and penalties are equated in conformity with the price analogy, some precautions will have to be taken. To begin with, one would need to account for the difference between trivial and non-trivial prices (or penalties). A trivial price (or penalty), as I said above, is a price that any agent holding certain preferences and resources might be willing to pay, even in conditions of less than perfect enforcement. But if a non-trivial price is, on the other hand, a price whose value no reasonable agent would be willing to afford, despite her preferences and resources (and in conditions of perfect enforcement), then there is probably something there that can give us the key to try to finally understand the distinctive nature of a properly *criminal* penalty.

In relation to the other crucial feature, Rawls'ss silence is even more revealing, for it shows in what sense an adequate account of *criminal* penalties cannot do without a certain conception of *criminal* offenses.According to the Kantian scheme, for instance –which we do not have to accept, of course –a crime is any transgression of public law that "makes someone who commits it unfit to be a citizen." (MS 6:331) In his own terms, what separates a criminal transgression of law from a non-criminal one (e.g. a civil transgression) isthat, whereas a criminal transgression endangers the commonwealth, a non-criminal one just endangers an individual person (*cf.* ibid.) As I said, we are

not compelled to accept such a scheme. Nevertheless, something like an explanation must be given of what it makes of a *simple* offense a *properly criminal* offense.

In my view, which is derived from what have been said in a previous section of this paper (*cf. supra*), the distinctive nature of criminal offenses has mainly to do with the kinds of personal rights and goods that are affected by them. Here I cannot enter into details, but in general I tend to think that the non-negotiable fundamental rights that constitute the heart of a person's autonomy are precisely those that a *criminal* justice system will want overwhelmingly to protect.[7] In this context, the expressive function of penalties becomes salient, for they will not merely purport to prevent future offenses or to restore a lost equilibrium, as would be the case, for example, if we were working within a corrective justice theory of tort law –according to which the plaintiff is supposed to be returned to her *status quo ante* (*cf.* Goldberg and Zipursky 2012: 25). Instead, penalties will mainly purport to communicate a message that something significantly wrong has been done. Morality, in any case, will have to enter the picture.[8]

To sum up, attending to these differences, it can be helpful to trace another distinction, namely that between what may be called *Trivial Systems of Punishment* (**TSP**) and *Non-trivial Systems of Punishment* (**NSP**). A **TSP** will be, in general, any system of punishment that satisfies the Rawlsian definition. In order to be a **TSP**, it suffices if a system reunites all the features captured in Rawls'ss definition of punishment, including a proportionality test based on a utilitarian preference-centered account of penalties. And, as can be inferred, it is not necessary for it to embrace a Feinberg-inspired condition, which demands the presence of an expressive element in every kind of penalty. Furthermore, a **TSP** is, in principle, easily compatible with any penalties enforcement lawlike systemthat may have as its central function to rectify or correct harms. In this sense, a contract law enforcement system as well as a tort law enforcement system can both be considered to fit in perfectly well with the characteristic normative structure of **TSP**.[9]

On the other hand, a **NSP** will be, by all means, any system of punishment satisfying the Rawlsian definition, but with a double *surplus-condition*: a) it must include a proportionality test that pairs off penalties and offenses on a non-merely preferential account of penalties, be it utilitarian or

[7] Here I am in broad agreement with Braithwaite and Pettit's position, though the concept they use is not 'autonomy' but 'dominion'. As they say, the evil associated with central cases of crime implies the invasion of dominion, which is constituted by an agent's person, province and property. Crimes such as murder or rape trespass against the victim's person; crimes such as kidnap or harassment trespass against the victim's province; and crimes such as burglary or theft trespass against the victim's property (*cf.* 1990: 69). What specific kinds of trespasses should be criminalized and how they should be so, however, raise quite different and intriguing questions indeed. Braithwaite and Pettit reasonably believe that in order to promote dominion – or to minimize its invasion – "criminalization is not the only way of inhibiting behavior; indeed it is probably the most clumsy and intrusive means available to the state" (1990: 71).

[8] On Goldberg and Zipursky's philosophical account of tort law, it is *wrongs* and not *simple harms* or *losses* what tort law is mainly about, where a wrong implies a "breach of a duty not to harm [...] through careless conduct." (2012: 35) At face value, it seems as if there was nothing really significant in this theory to account for the distinctive wrongness of criminal offenses. Indeed, as the authors explicitly recognized at the end of their paper, one of the main functions of a system of tort law is to further those values that morality tends to regard as important, "such as liberty of action, security against injury, attention to the interests of fellow-citizens, and the like." (ibid: 37) The furtherance and protection of *these values*, as they call them, seem to coincide, precisely, with the aim I have assigned to a criminal justice system. Is this not an unwelcome consequence for a view like my own? The key element to understand why it is not, I think, is provided by the same authors, who say that the distinctive characteristic of a tort law system is to further these ends *by empowering individuals* (not government officials) to hold wrongful injurers responsible *to them* (*cf.* ibid: 27; 37). In contrast, a criminal justice system will try to accomplish this task even when the victims were not able to see that they were injured. It is the state, and not individuals, that must decide when a wrong was committed and when a value needs to be reassured. The difference between the two kinds of wrongs, then, is quite obvious. Behind the tort law conception of wrongs, there lays a defeasible conception of values, which means that unless individuals decide to seek redress, they will not be fulfilled (*cf.* ibid: 37). On the contrary, behind the criminal law conception of wrongs, there lays a non-defeasible conception of values, which means that there are some values that cannot be violated, no matter what.

[9] Imagine that a criminal justice system is built in such a way that the penalties it distributes among citizens can be afforded with the help of insurances. In principle, a system like this seems perfectly capable of meeting each and every single one of the conditions enumerated by Rawls in his definition of punishment. However, even in a system like this, not every claim would be of a kind that could be met by insurance (*cf.* Williams 1993: 70). What are we supposed to say, for example, about those claims looking for an apology? What about repentance? Can we try to live within a system that ignores these social functions?

not[10]; and b) it must also embrace a Feinberg-inspired condition, which here means that the expressive element of condemnation be not simply directed to the contingent unreasonableness of the agent but to her moral status. By the way, in order to avoid confusion it must be noticedthat it is not only our current systems of punishment – as we actually know them – those that will fit in with a **NSP** structure. Assuming that we adopt a philosophical approach to tort law *à la* Goldberg-Zipursky (e.g. a civil recourse theory) (*cf.* 2012), then many of our civil law enforcement systems will also fit in with thetypical **NSP**'s normative structure. For our present purposes, such a disappointing implication can be regarded as totally innocuous.[11]

Now, if we come back to our original question and ask again how the Kantian dilemma can be solved, the answer seems quite obvious. Within a **TSP**, it can perfectly be the case that an offender may want (will/consent) to be punished, depending on factors such as her preferences, her resources and the enforcement characteristics of the particular criminal justice system to which she might belong. This is what I meant to call the easy way out of the dilemma.[12] By contrast, within a **NSP**, as I tried to argue, things look radically different. Indeed, within a system like this, a promising way out of the dilemma seems at least rather problematic to foresee. Notwithstanding these impediments, there are at least two theories that tried to defendthe view that the best way to justify the infliction of punishment, where punishment must be interpreted as operating within a **NSP**'s sort of system –even in conditions of partial enforcement, is by appealing to the offender's consent. In the next part I will try to assess these approaches in a critical way, showing that the notion of consent that they work with is conceptually spurious and justificatory inert, for it has nothing to do with the notion of consent that typically plays a justifying role, for instance, in a contract-law context.

III. The Consensual Way Out

Deterrent justifications of punishment have exerted a wide and deep influence among philosophers. In Rawls'ss case, as we saw above, that influence was incontestable, especially in his younger years. In more recent days, and in spite of the several objections that were raised against it, deterrence could manage to keep its philosophical appeal. Probably the most serious of these objections has an undeniable Kantian flavor, and it was raised even by those who defended a consequentialist justification of punishment but at the same time feared that it could infringe the moral imperative prescribing that we should never treat people *merely* as a means.Especially for some of those deterrent theorists who share this fear but do not dare to fall in the hands of retributivism, which was Kant's choice, a *consensual justification* of punishment was the right way to go.

In the present part, I will briefly analyze two of these justificatory strategies: Claire Finkelstein's and Carlos Nino's. What would lie behind them both is a *fairness owed to consent*'s approach to social morality (*cf.* Honderich 2006), according to which once an individual has given

[10] *Cf.* Williams 1973, where the author denies the possibility of a non-preferential account of practical justification conceived in a utilitarian vein. According to me, utilitarianism can do much more than that but it is not my intention in this paper to further argue in this direction.

[11] For further details on this discussion, seefootnote 7 above. On the other hand, Braithwaite and Pettit seem to have in mind a similar **NSP** model of a criminal justice system when they analyze the different sorts of sentences that should be permitted or enjoined (see 1990, Chapter 6, Section 2). In accordance with their previous distinction relative to what counts as an agent's dominion (*cf. supra*, footnote 4), Braithwaite and Pettit write that "the punishments imposed by courts can be neatly divided into three kinds, turning on our earlier distinction between an agent's person, his province, and his property" (1990: 102). Of the three kinds of penalties, however, the last one (in the form of fines, restitution, or seizure of assets) is also compatible with a **TSP** model.

[12] In a comment made on a previous version of this paper, José Juan Moreso suggested that from the novels of the 18th century we know the behavior of people disposed to commit minor crimes with a view to go to jail and spend the winter period there. In 2010, *The Guardian* newspaper said that a "fifth of homeless people have committed *imprisonable offences* to spend a night in cells" and "crime plays a big part in rough sleepers lives". Nearly 30 % of them admitted that they had committed a "minor crime such as shoplifting or anti-social behavior in the hope of being taken into custody for the night". Therefore, it is undeniable that there are cases in which people commit crimes with the only intention to be punished. However, my contention here is that although this kind of information can perfectly accommodate a **TSP**, it cannot accommodate a **NSP** so easily, which is what I try to argue in the paper. I would like to thank José Juan Moreso for raising this as well as other important issues.

ISSN-e: 2445-0669
DOI 10.7203/REK.2.1.8813

his/her consent to be deprived of some good, any objection to such a burdensome state of affairs will remain, *prima facie*, out of place. Whether punishment can be one of these burdensome states of affairs, however, is what I would like to challenge from now on. Of course, neither Nino nor Finkelstein seem to be suggesting that in order to offer a prominent account of justification, the concept of consent must be the expression of an individual's real will or anything like it. Quite the contrary, they both conceive its justificatory function as operating at a different level. Here, my main purpose will consist in showing that, even if we take these attenuating moves into consideration, consent to punishment will remain looking in our eyes as strange, nonsensical and counterintuitive as it must have looked to people over 5000 years ago.

III.1. Finkelstein's Consensual Proposal. Why not?

As the partial deterrent theorist she recognizesherself to be, Claire Finkelstein sees a crucial defect in any purely deterrent account of punishment. Resting on Rawls'ss terminology, she attacks any account like this by saying that it violates the "*no traveling across persons*restriction", which establishes the impermissibility of making someone suffer in order to prevent some other agents from suffering (2005: 212). Such an objection has been so extensive and persuasively defended in the literature that it is not my intention here to sound reiterative (see, for instance, Rawls 1971: 27; Nino 1983: 183-84; Finkelstein 2005). It is much more important for our present purposes to assess Finkelstein's own proposal to overcome it. Labeled by her as a *contractarian approach* or, alternatively, as a *consent-based approach*to punishment, what it maintains is that "the way in which the aim of deterrence is incorporated into punishment theory is not premised on total, or even average, social utility, but on the assent of each individual to the scheme by which such deterrent ends are pursued" (2010: 320).

In order to find out what this precisely means, we need to take a look into her main philosophical assumptions. Finkelstein begins again with the Rawlsian idea that society is itself "a cooperative venture for mutual advantage" (Rawls 1971: 4) and reasonably interprets it as implying that "society is the product of agreement among rational agents who see themselves as advantaged under the terms of social interaction, using as a baseline how they would fare in its absence" (Finkelstein 2005: 214). Immediately after that, she adds that one might even depart from Rawls and treat his idea "as something in the nature of a *requirement* for the basic institutions and practices that make up the fabric of social interaction" (ibid.). And then she asks: what basic institutions are supposed to be agreed upon by rational agents that are meant to decide when they will be better off? Amongst the basic institutions that agents will probably select,there is, of course, the institution of punishment, for it is reasonable to assume that each and every single one of them will be left better off than they would be in its absence (*cf.* ibid.: 215).

So far, everything seems to be in perfect order. Given that without an institution like punishment members of a social order would have no way of ensuring compliance to the terms of the agreement, punishment passes the benefit test and thus becomes justified (*cf.* ibid.). This compliance, it must be noticed, would not be possible if punishment did not make its job through the deterrent function that its threat supposes. However, what about the "traveling across persons" objection? In Finkelstein's view, the consensual nature of punishment defeats the concern because "each party to the social contract agrees that he will submit himself to punishment in the event that he would violate the conditions of the social contract". Indeed, she emphasizes, "it is this self-imposed threat that he offers to his fellows as his assurance that he will not defect", to conclude that "punishment itself is legitimate to inflict, not because it deters others, but because it has already been consented by the offender himself" (ibid: 217).

As I said in the introduction to this part, my reply to Finkelstein's proposal will be very brief. In particular, the only thing that really worries me is related to the kind of consent chosen to do

the justificatory work. There is one line of criticism that is not going to be here of my direct concern but that is nonetheless worth mentioning. Finkelstein's notion of consent is, in a way, exactly like the notion Rawls employs in the Original Position's argument, namely a *hypothetical* notion of consent. It has been convincingly argued that such a notion, taken in certain contexts, becomes absolutely unfair when is used to justify the imposition of significant losses on *real* people, disregarding their tastes, preferences, social positions, and the like (*cf.* Scanlon 1982: 122-33; Wannberg 2003: 32). This line of criticism, however, would apply to Finkelstein's approach if and only if it pretended to conflate into one single notion the will of the offender *as a social contractor* and the will of the offender *qua offender*. I think that a charitable reading of her position can avoid this undesirable consequence, making it plain that when she refers to the consent of the offender, the reference is given by his will as a social contractor –full stop.

The problem is that once we accept this charitable reading, we are unable to provide an answer to the "travelling across persons" objection, which is, as a matter of fact, the very motive that triggered her own ameliorating proposal. Indeed, the notion of consent that Finkelstein develops to justify punishment is in principle replaceable by any act of the will of a hypothetical contractor trying to make up his mind in an original social contract. That very act, nevertheless, exhibits no differences whatsoever if we compare it to the act of will that can lead a rational agent to choose and be guided by institutions such as tort law or contract law. For the sake of the argument, let us just focus on this second kind of institution. When someone signs a particular contract and accepts a given authority to enforce it, what is he supposed to be doing? What is his will's content? Are we supposed to believe that he is expressing consent to be guided by a social institution such as the one that made it possible to sign not only this but any other possible contract? Maybe that is true, but in no sense does it offer the slightest clueto help us start justifying the enforcement of that specific contract on that particular occasion.

With the notion of consent that happens to be at stake in Finkelstein's justificatory approach the problem is almost the same. In order to determine how and why that particular criminal offense has to be punished, the consenting will of the offender *as a social contractor* is not relevant at all.To put it in Kantian terms, it is his *phenomenological* will that we are interested in, not his *noumenal* one. In fact, that is precisely the argument that Kant offers in *MS*, when he says that one thing is to will to be punished and quite another is to will *a punishable* action. Thus:

> As a colegislator in dictating the *penal law,* I cannot possibly be the same person who, as a subject, is punished in accordance with the law; for as one who is punished, namely as a criminal, I cannot possibly have a voice in legislation (the legislator is holy). Consequently, when I draw up a penal law against myself as a criminal, it is pure reason in me *(homo noumenon),* legislating with regard to rights, which subjects me, as someone capable of crime and so as another person *(homo phaenomenon),* to the penal law, together with all others in a civil union. [...] For, if the authorization to punish had to be based on the offender's *promise,* on his *willing* to let himself be punished, it would also have to be left to him to find himself punishable and the criminal would be his own judge (MS 6:335).

Finkelstein's approach, unfortunately, has proven to be too ideal to provide an answer to the Kantian dilemma. Naturally, since Finkelstein's intention has never been to accomplish such a task, there is no reason to blame her for not doing so. The possible success of her approach can only be measured against the benchmark of herown attempt to meet the objection against the purely deterrent account. It was in relation to that attempt that her notion of consent was introduced. However, even if we are guided by that criterion, her proposal does not improve, for it is her notion of consentthe one that is conceptually spurious and justificatory inert. In what comes next we will try to see whether Nino's approach is better suited to deal with the '*travelling across persons'* objection' problem, providing a more robust notion of consent.

III.2. Nino's Consensual Proposal. Why not?

In *Towards a General Strategy for Criminal Law Adjudication* (TGS) (1977) as well as in a plurality of essays coping with different criminal topics (the death penalty, drug abuse, proportionality between harm threatened and harm averted, self-defense, etc.), but mainly in "A Consensual Theory of Punishment" (1983), Nino has tried to offer a twofold justificatory approach to the practice of punishment. On the one hand, as I said above, Nino openly confesses his keenness on the typical deterrent justification addressing consequential considerations of social protection. Yet, aware at the same time that a purely utilitarian justification of punishment could be objected on behalf of the Kantian worry that we should never treat people *merely* as a means, Nino offers, on the other hand, a consensual type of justification whose main desideratum goes exactly as follows: "The fact that the individual has freely consented to make himself liable of punishment (by performing a voluntary act with the knowledge that the relinquishment of his immunity is a necessary consequence of it) provides a *prima facie* moral justification for exercising the correlative legal power of punishing him" (Nino 1983: 299)

As might be seen, two are the conditions Nino seems to have signaled as shaping the *definiens* of consent: (1) a voluntary act (which here means, essentially, a free and not coerced act) *plus* (2) a certain knowledge or foresight of the normative consequences that it leads up to (which in the case of punishment means, as seen earlier, a relinquishment of immunity to punishment, 'which is to say to the gaining of a power by officers of the society') (Honderich 2006: 50). Moreover, were these not the defining conditions of consent, it wouldn't have made sense that Nino had chosen to appeal in TGS to an analogy: "It is a matter of positive law –he wrote as a way of illustration –that (...) taking something off the shelf of a supermarket involves the obligation to pay the price, that accepting to travel with a drunk driver means (according to some opinions) to waive the right to compensation in case of accident, etc." (Nino 1980: 231)

Based on this analogy, some interpreters noticed that the concept of consent that Nino imagined to play a role in justifying punishment was the notion of 'tacit', 'alleged' or '*ex actionem*' consent (*cf.*Boonin 2008: 164; Malamud Goti 2008: 227-255; for a contrary view, see Imbrisevic 2010), a notion that, as we all know from Locke, raises a number of semantic problems whose severity is not to be taken lightly (*Second Treatise on Government*, §119; see also Simmons 1993: 83-87; Lloyd Thomas 1995: 39; Boonin 2008; Imbrisevic 2010: 213-214). These specific semantic problems, however, will not be here of my concern, for I am manly interested in the analogy that Nino draws between contract law (and tort law), on the one side, and criminal law, on the other. Based on it, consent to punishment is paralleled, for example, to the assumption of risk that takes place whenever one gets involved into dangerous activities which may require our waiving the right of compensation in case of an accident. Thus, assuming that the criminal act was voluntary and made after having the chance to foresee the normative consequences it may lead up to, punishment for that act is supposed to be *objectively* consented.

At first glance, it gives the impression as if Nino's notion of consent were conceptually stronger than Finkelstein's one. Hence, whereas Finkelstein's approach failed because it merely seemed to rest on the *noumenal* will of the offender, in that sense Nino's approach seems to be better suited to do the justificatory work. To say it again using Kant's terminology, his approach seems to have been conceived not for the *noumenal* but for the *phenomenological* world instead. And taking into consideration that it seems to have been conceived to operate withina **NSP** in conditions of less than perfect enforcement, Nino's justificatory strategy appears to be even more ambitious from a practical point of view. Unlike Rawls, Nino is perfectly aware that punishment cannot be equated with a simple monetary fine. Punishment implies the infliction of great suffering on the person of the

wrongdoer, which is why it demands a special kind of justification. In any case, the problem with it would lie, as I will try to show as briefly as I can, in the analogy on which it rests. My humble purpose until the end of the section will be to debunk Nino's account by challenging the nuances of his analogy.

Recall again what it was said in a previous section regarding the analogy between penalties and prices (*cf. supra*). *Mutatis mutandis*, something similar is valid in the present context. Before consent can find its place in a general justificatory strategy of punishment, a lot of extra work needs to be done to allow the types of actions typically involved in each area of law run in parallel. Nonetheless, for reasons that have mainly to do with the characteristic elements that define the nature of each activity, such a work will prove in the end to be hopeless. Just to take a simple case by way of illustration, think of those dangerous activities, such as undertaking a life-saving surgery or the practice of extreme sports. While the risk involved in those activities is a non-dissociable part of their nature, the greatest risk associated to criminal activities – e.g. punishment – usually depends on what the state does as the agency that claims the monopoly of violence. Indeed, while it wouldn't make any sense at all to get involved in a naturally risky activity without *somehow accepting* to pay its costs, it makes perfect sense to commit a crime – or an action deemed *criminal* – without accepting to bear its merely contingent consequences.

Our ***(PRS)*** practical syllogism can be of help here too. If we say that agent ***X*** wants (wills) to practice parachuting ***(P)*** (S1) and that it is highly unlikely that practicing parachuting ***(P)*** can be accomplished without implying a certain risk to be seriously injured as a consequence of a failure in the parachute mechanism ***(Q)*** (S2), therefore, if conditions (S3…S6) are met, we can arrive at the conclusion that agent ***X*** will consent to assume that very risk ***(Q)*** (C). In such a case, of course, *consenting*has a lesser practical status than *wanting* or *willing*, which is the proper status that comes withsome of our subjective attitudes in regard to certain affordable prices (see again our practicing yoga example, *cf. supra*). However, a conceptual difference like this, rather than invalidating the analogy, makes it even more interesting, for it paves the way towards achieving a more compelling case for consent to punishment. It must be noticed by the way that, according to Nino, it is not necessary that a positive subjective attitude be found in the offender for consent to punishment to be possible. Quite the contrary, as Imbrisevic remarks, "I can consent reluctantly – and even without hiding my reluctance" (2013:115); moreover, I can consent less than wholeheartedly, for example with *grudging acquiescence* or even out of indifference (Barnett 1992: 866). In none of these cases, however, the phenomenological character of the subjective attitude would turn it into less consensual.

With these clarifications in view, let us try to apply our syllogism to Nino's account. What would the first premise (S1) be? What isthe ***P*** that ***X***, the offender, may be said to will? Assuming that it is the state of affairs that is brought about as a consequence of his implication in a criminal activity, like the money he could get from assaulting a bank, the conclusiveness of the syllogism will depend on what we say next to interpret ***Q***. For the sake of the argument, let it be the risk of being caught by the police authorities and then prosecuted by the public officials of the state. If it is highly unlikely that ***P*** be accomplished without implying ***Q*** (S2), then, as in the case of practicing parachuting, it is reasonable to presume that the willingness of ***P*** will be tied to the risk's materialization degree. When the risk is too high and can be known in advance (for example, when the parachutist knows that the parachute mechanism is not well-designed or whatever / when the offender knows that the enforcement conditions are almost perfect), the willingness of ***P*** (let us say, the enjoyment of free falling / the enjoyment of money) might be seriously distorted.[13]

[13] As one of my anonymous referees wisely noticed, there seems to be an important difference between saying of an agent who decides to practice an extreme sport that she assumes the consequences that may be generated as a product of it, and saying that what she is actually doing is to want certain necessary means to her end, which is actually the Kantian claimbehind the practical syllogism. Recall, however, that whenI refer to the parachute case, for instance, I do not take for granted that what the agent wants

In contrast, let us suppose now that the risk *(r)* is too low and can be known in advance with an important degree of certainty. Reasonably, agent ***X*** may still want ***P*** and can bet for her action vigorously, with the hope that *(r)* never materializes. If, for whatever reason, *(r)* finally materializes and the state of affairs ***P*** is not realized, there are some cases in which we would still be willing to say of agent ***X*** that she wanted to doexactly what she did, even when that produced a different outcome. Regarding the parachute case this is perfectly clear. Once we are certain that the parachute failed because it was in the nature of its mechanism to fail and there was nothing within ***X***'s power to prevent that failure, it is perfectly legitimate to affirm that ***X*** has no motives to blame herself for having acted as she did. Indeed, that is the reason why we might even dare to say that ***M*** [namely the materialization of risk *(r)*], was consented by ***X***, although ***M*** implies ¬***P***.

Regarding the criminal case, however, things are not that simple. In order to be able to say of an offender ***X*** that he has no motives to blame himself for ending up in prison, which is what ***M*** generally implies here, there must be granted that ***X*** didn't do anything wrong to give rise to the unwelcome outcome.[14] However, when it comes to determine the risks that are materialized when a criminal act is committed, in a vast amount of occasions it would look as if the offender could always have done something different to avoid being caught. This impression is confirmedevery time we get to the conclusion that the offender behaved in a careless way, or even out of ignorance of the possible consequences of his acts. But not everything has to be the offender's fault after all. In contrast to the parachute example, in criminal contexts it is sometimes very hard to tell what the risks of being caught are for breaking the law. Our criminal systems tend to be inherently stochastic in their functioning, which means that no one (perhaps with the exception of a few authorities) is ever capable of determining with precision what will happen.In those contexts, it would seem rather unnatural to say of an offender that he has consented to go to prison, even if going to prison was one of the possible implications of his actions.

Compared to the notion of consent that was operating in the original practical syllogism ***(PRS)***, the notion of consent we have arrived at in the end looks indeed very peculiar, to say the least. Consent is here equated with the absence of reproach for how things unfortunately turned out to be in spite of my preferences and actions. Such a notion, of course, should not be mistaken for conformity with respect to a certain state of affairs, for there is nothing in the resulting state of affairs that actually conforms to my desires. But given that, as Williams would say it, "I acted and deliberated as well as I could, and it is sad that it turned out that way" (1993: 69), there certainly is a kind of conformity with myself, a kind of irremediable and peaceful acceptance of the circumstances, that will not always be present in the post-factum evaluative stance of the punished offender. Quite the contrary, even after doing his wrong, the offender will probably struggle to evade punishment by defending himself in a fair trial according to the due process of law. And even after being declared guilty, if he were given the chance to escape, he would probably take it (*cf.*Honderich 2006: 51).

Nino's consensual theory of punishment is more than aware of that peculiarity. For this reason Nino has repeated many times how his notion of consent was supposed to be interpreted. When someone commits a criminal act, what he shall be taken to be consenting to *through* the act *per se*is a relinquishment of immunity to punishment, which only means that he is renouncing to the righthe would otherwise have to demand from the officers of the society a kind of redress for

---

are the consequences of her actions, which might be sometimes almost impossible to predict. Here we should be very careful to draw a line of separation between the predictable risks associated with an action and the real consequences that are materialized afterwards. My argument rests on the former, not on the latter. Of course, the risks assumed when doing X with a given end in view do not coincide with the means to further that very end, which is Kant's point. But that is precisely the reason why, when presenting ***(PRS)*** in the original version of this article, I should have writtenthat it was*inspired* by the Kantian principle, instead of *governed* by it –as I actually did (*cf. supra*, 2.2, especially footnote 3above). I would like to thank my reviewer for noticing these differences, which allowed me to improve this presentation.

[14] Blame, as might have been guessed, is not being used here as a moral notion. For a more comprehensive notion of blame, see for example Williams 1995 and Sher 2001.

ISSN-e: 2445-0669
DOI 10.7203/REK.2.1.8813

inflicting him that sort of wrong. If we assume the validity of this explanation, it would give us a way of justifying the loss of that particular right. The offender justifiably loses a right because he has consented to lose it. However, as Honderich notices, what seems to be at work here is "the offender consenting in a secondary sense to a necessary condition of punishment, and not consenting, in any sense, to his punishment" (2006: 52).

Consent to punishment implies, to be sure, quite a different loss, a kind of loss more stringent from the moral point of view, and it is this loss the one that still begs for justification. In a consequentialist approach like Nino's, it is out of question how the act of inflicting it becomes justified. Were it not for the fact that a loss like the curtailment of the offender's bodily movement's freedom, for example, succeeded in promoting a goal of social protection, under no concept would that be permitted. But, again, this implies treating the offender as a mere means, which is what Nino's justificatory strategy sought to avoid from the beginning. Therefore, the challenge remains: until the offender's consent can prove to be directed towards punishment itself and not towards a different substitute (e.g. a relinquishment of immunity), consent to punishment will look as counterintuitive and strange as it certainly looked to Kant and many others before and after him.

IV. Is it Possible a Way Out of the Dilemma?

From the beginning, the main goal that the present paper embraced was to challenge the idea captured in Kant's claim that it is impossible to will to be punished. Discarded what I called the easy way out of the dilemma, as well as the consensual way out that derives from Finkelstein's and Nino's justificatory approaches to punishment, it is now time to give a straight and final answer. Without further ado, I would like to say that it is perfectly possible to will to be punished, even if that willingness is not able to play any *justifying* role in a justificatory theory of punishment. Two, then, are the main commitments of this final part: on the one hand, to assess how it is possible to will to be punished (4.1); and, on the other, to briefly describe what is the exact place that such a willingness can occupy within a more general approach on the justification of punishment (4.2).

IV.1. What does 'Consent to Punishment' Really Mean?

In my view, the main problem that all the attempts analyzed in this paper face to find out a convincing way out of the dilemma is that they are lookingin the wrong place. They all seem somehow to suppose that the only way to credit for the willingness of the wrongdoer to his punishment is by trying to determine whetherthe very criminal act expresses a subjective attitude that may transpire such a flavor. These attempts, however, often ignore that when people commit crimes, they tend to behave more or less in the way described by the *Book of Job*. For offenders, "midnight is their morning"; to get away with it, "they make friends with the terror of darkness". In order to reverse the tendency, what we need to keep in mind is that whatever one may feel towards punishment, one may get to feel that at a later stage, for example during the judicial procedure that is judging a suspect or even after a conviction sentence is passed against him.

In *Punishment, Communication and Community* (2003), Anthony Duff did a great work to clarify this idea, which owes much of its force to what I referred above as the second defining feature of punishment, neither acknowledged in a scheme like the Rawlsian one (cf. *supra*), nor in a scheme like Nino's.[15] According to Duff, there is no way of conceiving punishment independently of its

[15] Here I cannot argue as I should to support this claim. Instead, let me just offer a general remark. Even if Nino accepts the expressive element as one of the defining features of punishment (*cf.* 1980: 205), it was never clear whether that element finally managed to find a place in his theory. For instance, Nino rejected the blaming function of punishment, for considering that the notion of blame necessarily commits us to a perfectionist conception of the person. In ["Datoseliminadosparafavorecer la anonimidad del envío"], I extensively dealt with this issue, trying to demonstrate that it is perfectly possible to conceive a notion of blame fully consistent with liberalism.

ISSN-e: 2445-0669
DOI 10.7203/REK.2.1.8813

communicative and expressive functions. Punishment purports a message to the individual whom it is applied to, as well as to the liberal and democratic community which witnesses it. Hence, as a communicative entity, punishment must attempt to persuade a criminal to repent her/his crime and "to accept the values that condemn it", even if this persuasion can only take place in a forcefully institutionalized context. Yet, this forceful element is justified when punishment addresses "the offender as a member of the liberal polity whose autonomy must, like that of any citizen, be respected". And punishment manages perfectly well to do this for it is, "like any exercise in rational communication, necessarily fallible: not because it might in fact fail to persuade him, but because it must itself leave that possibility open" (2003: 122). In the end, that is why justified punishment may be conceived as an authentic consensual enterprise, i. e., because by appealing to the conscience and understanding of the offender, embracing her/him as an autonomous and rational agent, it seeks for her/his consent as a necessary and final source of legitimization.[16]

But even if this alternative were foreclosed, I think that there is another plausible way to conceive an offender's consenting to his punishment. Conceptually, this way shares the focus of the approaches previously objected, for it takes as central the manifestation of consent that is produced *ex ante* the judicial procedure, when punishment is still a remote possibility. Nonetheless, in contrast to these approaches, I think that it can be preserved from some of their most dubious assumptions. Imagine, for instance, a case like this. Harry Fertig in Sol Yurick's novel (1966) killed seven workers of the Mercy Memorial Hospital, including well-known doctors and its administrator. The motive, as Fertig confessed for the first time to one of his captors, was that they were responsible for the death of his son, Stevie, who had been negligently treated a few months ago. "You see – Fertig says –, killing balanced it out, and my being caught closes the account. That's why I had to be caught" (1966: 38). Or recall, on the other hand, the case of Anders Behring Breivik, the right-wing extremist who was found guilty of murdering 77 people by the Oslo District Court. When confronted to the judging audience, Breivik did not only assume the authorship of the horrible murderers, demanding to be punished; neither was he even interested to show a minimum gesture of remorse.

Evidently, cases like these are extremely rare. However, once we are invited to trust in their realism, a mere preferential account of human rationality will soon reveal its insufficiency, inadequacy and incompleteness to understand their nature. For what these cases expose like no other is that, even under Non-trivial Systems of Punishment (**NSP**) in conditions of perfect enforcement (*cf. supra*), there could be human beings who might not be deterred to break the law. Under Trivial Systems of Punishment (**TSP**) in conditions of almost perfect enforcement, the natural account of the typical case will go approximately as follows: it is reasonable for agent ***X*** to break the law when the subjective costs that she will have to face for refraining to commit an offense are not only higher than the costs entailed by the corresponding penalties (after all, high-certainty regimes usually require low-severity penalties) (*cf.* Kahan 1997: 379) but even higher than the costs of trying to avoid being caught. But, of course, in cases such as Fertig's or Breivik's, the costs implied in avoiding being caught were by all means set apart from the pack.

Following again our practical syllogism, we can think of these individuals as people who want to kill other people ***(P)*** even taking for granted that ***P*** certainly implies 'being punished' ***(Q)***, but who, having figured out scenarios ***a*** [***P*** and ***Q***] and ***b*** [¬***P*** and ¬***Q***] as the two exclusive alternatives, are more than eager to choose, out of non-preferential reasons, scenario ***'a'*** over scenario ***'b'***. Here it is really important to emphasize the non-merely preferential character of those reasons, for it is much more than simple preferences what push them to make up their minds. Typically, most of these cases

[16] Alfredo Traps in Dürrenmatt's tale (e. g. *Die Panne*) is the perfect example to illustrate the whole point, although not Duff's particular position. At the first stage of the imaginative judicial procedure that is going to determine his presumed guilt, Traps is totally convinced of his innocence. However, as the procedure goes on, he begins to discover not only that he has brought off a murder but that the same murder merits admiration, astonishment and respect. In the end, he "makes claim to the murder as his and demands the prescribed punishment – death" (Morris 2001: 257).

involve individuals who do not care for their fortune (a life in prison, let us say) as much as they care for their moral selves or personal identities. They may feel that unless their personal values – honor, reputation, self-esteem, honesty, heroism, civility, social commitment and the like – are somehow vindicated or preserved, their integrity as agents will be in serious danger.[17]

Fertig, for example, is a clear sample in this sense. According to Yurick's description, he is neither a madman nor a murderer. He is "a victim who has risen up against corruption and man's humanity to man […] by offering his life up so that it may no longer be" (1966: 251). By doing what he did and accepting to be punished, but only by accepting to be punished, he experiences a second birth, a birth that, "unlike the first one which merely brings us into the world of man, […] cuts off from all mankind" (1966: 281). As I said, this is a perfect sample. But there may be others. A lover who kills his/her partner because s/he resents his/her infidelity can be a paradigmatic example (Kahan 2001: 177). Rawls'ss civil disobedient can be another example too, since he publicly breaks the law but expresses at the same time his fidelity to it "by the willingness to accept the legal consequences" of his conduct. As Rawls put it, this is the price to be paid to convince our fellow citizens that "our actions have, in our careful considered view, a sufficient moral basis in the political convictions of the community" (Rawls 1969: 182). Assuming that they are credible cases, do not they represent authentic ways of tacitly or even explicitly consenting to punishment? I truly think that they do.

One additional feature is worth mentioning before turning the page. In section 3.2 above, we have seen that an agent who embarks in a course of action whose consequences happen to contravene her desires has no motives to blame herself for her misfortune as long as there was nothing within her power to prevent it. Of course, she might regret for how things turned out to be in the end; but this is quite a different subjective attitude (*cf.* Williams 1993: 69). Regarding non-law abiding conduct, we have also seen that for a non-law abiding agent it might be much harder to escape personal reproach, especially when the criminal justice system is organized in such a way that there is no chance to predict with precision what to do to avoid being punished. Finally, I tried to show that when an agent blames herself in such a way for the resulting outcome, any attempt to read in it a consensual elementwould be pointless. This sort of personal reproach, however, shall not be confused with the moral reproach that the agent might want to direct against herself for breaking the law, assuming that she is not the kind of amoral agent that used to be the topic of many philosophical discussions in the past. Nor shall it be confused, for the same reason, with the moral reproach usually purported by punishment itself.

Now, the additional characteristic feature of consent to punishment that I would like to mention can be interpreted as a corollary derived from these considerations. When consent is absent, as is generally the case, for instance, in those systems of punishment operating in conditions of less than perfect enforcement, in principle it would be much easier for the blaming function of punishment to get a grip. This is not difficult to see. Compare two non-law abiding agents: Fertig and Vertig. Whereas Fertigcommits a crime and confesses his guilt, Vertig commits the same kind of crime but tries to conceal it. Whose conscience looks more opprobrious? Anyone can see that it is the second agent's conscience the onethat punishment, accomplishing its moral expressive function,will be mainly interested to be directed to. By contrast, when consent is present, as Fertig's case clearly exemplifies, punishment might experience some expressive limitations. Obviously, that's not impediment to try to condemn every act that turns away from the law, especially when they seem to be the manifestation of an individual's non-sharable idiosyncratic values. However, given that in consensual kinds of crimes it is the very wrongdoer the one who has accepted to be punished, it is likely that the expressive function of punishment be directed mostly to those members of society that

[17] For more extensive remarks on this point (especially regarding moral and immoral actions) see Cohen 1996, Geuss 1996 and Frankfurt 1998.

contemplate the crime and its punishment from the outside, perhaps as mere impartial spectators, perhaps as their direct or indirect victims.[18]

IV.2. Molding an Alternative Consensual Approach

If the reasons presented so far are credited with some plausibility, then consent to punishment would turn out to be possible even under Non-trivial Systems that distribute penalties in accordance with a non-merely preferential account of human rationality. Nonetheless, in order to infer from this premise the desired conclusion, namely that consent is what justifies the imposition of punishment on an individual, absolutely nothing of what have been said so far will do.

Furthermore, assuming that agents such as Fertig and Breivik did both express authentically consensual attitudes to be punished and although consent may find in their cases a justifying function – which, of course, need not be the case at all –, there is almost nothing there that can be useful to cope with the vast universe of offenders, unwilling to pay for what they have done. In contexts where it really makes sense to talk about consent and consent plays a justificatory role, we tend to believe that it does, to quote David Boonin, if and only if it "can be overridden by an explicit declaration of the contrary" (2008: 164). But once we reach this definition, we easily get to appreciate its fatal implications for the criminal context. For, as Boonin remarks again, if the application of punishment must be mediated by an individuals' consent, "any offender could free himself from liability to punishment merely by announcing that in doing a certain act he did not intend to consent to liability to punishment" (ibid: 145).

Therefore, what we need to do to know what amount of special attention must be given to the concept of 'consent to punishment' is to start acknowledging thatthe specific role that certain attempts to justify punishment have unsuccessfully conceded to it in the past shall not lead us to believe that there is no other alternative. In my view, the notion of consent is perfectly capable of playing a role in a general approach to evaluate the practice of punishment, even if that role is not itself a justifying one. At the beginning of the last section (4.1) I mentioned Antony Duff's appeal to this concept in an attempt to show how it is supposed to intervene in a *post factum* stage of a criminal law adjudication strategy, being useful to morally guide officers who run the execution of punishment. Instead, the problem thatconcerns me here is closer to the other notion of consent, the one that is at stake in criminal actions such as those that I tried to cope with above (*cf. supra* 4.1). The following lines are meant to shed some light on this problem.

One of the most difficult tasks that any comprehensive approach to justify punishment must face even today is how to match crimes with punishments. According to retributivism, *lextalionis* was the simplest way to accomplish this task, by insisting on the idea that offenders deserve to experience the suffering they inflicted on their victims. However, such an answer leaves us with empty hands when taken too literally, for "no one would advocate raping rapists, assaulting assailants, or burgling the homes of burglars" (Finkelstein 2010: 212). Probably, an assaulting penalty for an assailant will manage to deter many would-be assailants. Nonetheless, as the liberal and democratic citizens that we proudly claim to be, no one would even dare to come up with a proposition like that. But which are the penalties that must be morally admitted?

[18] In "Communication, Expression, and the Justification of Punishment", Andy Engen distinguishes between the communicative and the expressive functions of punishment. To fulfill its *communicative* function, the moral message of punishment must always be accepted in the end by the wrongdoer, who must show repentance and a disposition to reform her future behavior (*cf.* Duff 2003: 91-113). However, what if we have enough compelling reasons to believe that, in spite of what we do, the wrongdoer will not be willing to accept the moral condemnation that punishment purports? The objection of *The Unreceptive Wrongdoer*, which is the placeholder Engen uses to refer to "possible cases in which the punished does not respond in the desired way to the message punishment sends" (2014: 301), overrides communicative theories of punishment, especially those proposed by Hampton and Duff (*cf.* 2014: 300). For its part, punishment fulfills its *expressive* function even when it is not accepted by the punished. Engen writes in line with Feinberg and Scanlon: "Punishment that expresses condemnation, then, plays the important social role of affirming the rights of crime's victims and this could be among those things that crime victims are owed by the state" (2014: 305). In Fertig's case and cases alike, therefore, it seems to be precisely this expressive function the one that punishment is still supposed to fulfill.

In the last years, a bunch of compelling theories have been displayed on the screen. Restorative approaches to criminal justice, for instance, have made notable efforts to conceive alternative ways to deal with criminality, proposing to replace the more traditional penalties-based accounts of our criminal justice systems for other novel accounts, more focused on the needs of victims and offenders (*cf.* Braithwaite 2002; Urban Walker 2006). Deterrent punishment as such, defined as the necessary infliction of suffering fostered by a goal of social protection, does not seem to have a comfortable place in these accounts, for they are interested in promoting repair and reconciliation between victims and offenders rather than in simply preventing the occurrence of future crimes. In more than one sense, those attempts are highly valuable and it is not my intention to criticize them here. However, as Allison Morris acknowledges (2002), how they are supposed to effect real change and toprevent recidivism are among the main questionsthey haven't yet provided a straight answer for.

The alternative approach to punishment that I would like to propose here does not pretend to have a voice to deal with these legitimate concerns. For a similar reason, neither does it seek to position itself as a competitor within the well-known universe of justificatory approaches to punishment already existent, such as retributivism, preventionism, utilitarianism or even consensualism. As became clear from the discussion above – especially regarding consensualism – it surely shares some of their commitments as well as it rejects some of their implications. A general justificatory approach to the institution of punishment, however, is what my approach cannot pretend to be. With greater humility, it just seeks to offer in its place a theoretical framework that allows us to understand exactly what can be expected from a system of punishmentconceived in wide preventionist terms.And what it has to say in this respect is that, even if we were capable enough to conceive a system of punishment operating in conditions of perfect enforcement, still there would be citizens who might want to break the law. Who would they be? Well, precisely those who happen to consent to their punishment.

If we had to put a label on it I would be found myself forced to say that my approach is *alternatively consensual*. In choosing to call it that way, however, we should be preserved from the wrong impression that it might give, as if it were somehow trying to compete with the traditional consensual approaches previously criticized. As a matter of fact, the tentative label *alternatively consensual*does only make sense if it happens to suggest where its main interest is being directed to.Here we should proceed with caution. When consent plays a normative role, oddly does it so by offering a final source of legitimation or the like. In "Government by Consent", Joseph Razwas really emphatics on this point. There he claimed in relation to our consensual attitudes towards our own governments: "Consenting to be ruled by someone expresses confidence in that person's ability to rule well. […] consent has an expressive value only where the conditions of legitimacy are satisfied or are nearly satisfied. […] Hence consent has an independent, but auxiliary and derivative, place as a source of legitimacy" (1994: 351-53).

In a similar vein, Ernesto Garzón Valdés once said that although the factual consent of autonomous citizens is neither a necessary nor a sufficient condition for the genesis of a political system's legitimacy, this does not make it irrelevant to determine the actual effectiveness of the system, for such a factual consent is a crucial element of stability (*cf.*Garzón Valdés 1990: 22-23). Because I tend to agree with these authors that consent is called upon to play mainly a derivative and expressive role, no general approach to punishment can make use of it as a fundamental or sufficient notion. That is why, in my scheme, consent exclusively behaves as signaling how well a given system of punishment – whose general justification is to be sighted somewhere else – might be performing in the application stage of its organizational structure.

With that end in view, our previous battery of distinctions can be very helpful again. Recall that we have distinguished between Trivial and Non-trivial Systems of Punishments (**TSP** and **NSP**

respectively). Now, in order to evaluate how well a **TSP** can perform in its application stage, the evaluative criterion can be provided by what we may call a Consensual Ideal of a Trivial System of Punishment in conditions of perfect enforcement (be it **CI-TSP**). Whether such a System is empirically conceivable is not relevant here, for we are only interested in the evaluative criterion it offers. Hence, a **CI-TSP** in conditions of perfect enforcement can be characterized as seeking the following main goals:

a) To deter those crimes that would be committed out of mere preferential motives if the trivial preferential motives brought in by punishment were not as unaffordable as they are for some people.

b) To expectonly those crimes committed out of mere preferential motives that can be affordable by some people in spite of the imaginable contravening effects thatthe actual application of punishmenthas over those motives.

Alternatively, in order to evaluate how well a **NSP** can perform in its application stage, the evaluative criterion can be provided by what we may call a Consensual Ideal of a Non-trivial System of Punishment in conditions of perfect enforcement (be it **CI-NSP**). Assuming that a **CI-NSP** can be comprehensively conceived as to include a typical **CI-TSP**, a sort of *mixed* **CI-NSP** in conditions of perfect enforcement willthen becharacterized as seeking the following main goals:

a) To deter those crimes that would be committed out of mere preferential motives if the trivial preferential motives brought in by punishment were not as unaffordable as they are for some people.

b) To deter those crimes that would be committed out of mere preferential motives if the non-preferential motives brought in by punishment were not as unaffordable as they are for some people.

c) To deter those crimes that would be committed out of non-preferential motives if the non-preferential motives brought in by punishment were not as unaffordable as they are for some people.

d) To only expect:

i) those crimes committed out of mere preferential motives that can be affordable by some people in spite of the imaginable contravening effects that the actual application of punishment has over those motives; and

ii) those crimes committed out of non-preferential motives that can be affordable by some people in spite of the imaginable contravening effects that the actual application of punishment has over those motives.

It is in contrast to these ideal models that our actual systems of punishment have to be consensually evaluated. This task can be done in more than one way. In a general way, it would be almost impossible to determine with precision the consensual status of a system, forsystems tend to be highly selective in enforcing most of the norms that are codified. Nonetheless, in a more particular way, that can be done more easily, for it is just a question of trying to assess from the whole universe of possible offenses codified in a given criminal legislation, which specific kinds will be less or more likely to flourish. So, for example, if it happens that a vast amount of *type-A* offenses are being punished but this does not bring about as a result a *type-A* offenses' reduction, it would be legitimate to assume that the particular system of punishment under analysisis being consensually applied, sincetheir addressees seem to be willing to pay for what they have done. Being that the case, it is also legitimate to infer that such a consensual system – at least in relation to those offenses – is not accomplishing at all a proper deterrent function.

If it happens, on the contrary, that the system proves to be quite effective in sanctioning *type-B* offenses, an accomplishment that is followed by a *type-B* offenses' reduction scenario in a later phase of the criminal-law application stage, it will be legitimate to assume that most of the *type-B* offenses to be committed from then onwards are going to be the unmistakable expression of the offenders' consent to be punished. How vast can be the universe of possible consensual offenders for a system – thus organized – not to collapse is quite a different matter, which must be determined within the boundaries of each concrete system of punishment. As is reasonable to suppose, that will vary depending on different factors, such as the gravity and intensity of the rights' deprivation that punishment purports; the individuals' average tolerance threshold; the beliefs, values and desires that are predominant in a social community; the way punishment is seen by those who suffer it as well as by victims and citizens in general; and so on and so forth.

Be it as it may, it must never be forgottenwhat the notion of consent is capable of doing.In company with the aforementioned ideal scenarios that are theoretically conceivable with its help, what the notion gives us is just one criterion among others to determine how well a given system of punishment performs in order to protect the social values and individual freedoms that might be endangered by criminality. That is precisely why, in conditions in which the rates of law enforcement are very low and most offenders succeed to go unpunished, it will be less likely forcriminal conductsto exhibit subjective attitudes such as the willingness of their authors to be punished. According to preventionism, a system of punishment with these characteristics is, typically, a system of punishment failing to accomplish its proper deterrent function. Meanwhile, according to the alternative consensual approach that I tried to present in this paper, it shall be seen as more than this. In fact, it can also be seenas a system that failed to give individual consent the expressive derivative function that is supposed to be inseparable of any legitimating account of human institutions.

Last but not least, the notion of consent represents *just one* criterion among others to evaluate a system of punishment's performance because it is not much what it can really do by itself. For instance, it is not allowed to determine the fairness of a system. Imagine a *1984's* social scenario in which surveillance is really intense: there is a police officer on every corner, there are hidden cameras transmitting alive what happens in every building of every neighborhood, electronic bracelets for tracking persons under house arrest, satellite surveillance, and the like (*cf.* Braithwaite and Pettit 1990: 109-110). Moreover, imagine also that prosecution policies are organized "on the basis of possible rather than probable guilt" (ibid: 44). Under these brave-new worldly conditions, it may be very unlikely for a rational individual to will to commit an offense without expecting to be punished. We may even dare to say that, under these conditions, offenders will tend to express their consent to be punished. Is such a system fair or remotely desirable? Doubtfully could it be so, unless penalties were really affordable by anyone. But if they were not and the only offenses taking place were those perpetrated by individuals on the basis of their non-preferential motives, such a system would probably satisfy the consensual ideal being at the same time totally unfair in regards to many of the values and liberties that are sacred to the immense majority of the population.

A Consensual Ideal of a Non-trivial System of Punishment **(CI-NSP)**, therefore, need not be after all an *ideal system of punishment*. Nonetheless, it would be more accurate to say that a system of punishment cannot embrace any ideal target worth going for, whichever that is (namely restitution, furtherance of dominion, incapacitation of offenders, crime prevention, minimization of harm, etc.),ignoring the personal motives that people might have to break the law. The alternative consensual approach that I brought into consideration simply encourages measuring a criminal justice system's performance without taking our eyes off the complex world of human motivations.

To see the point more clearly, recall Sol Yurick's novel once again. The American justice system did not fail to prevent Fertig's actions because the law enforcement's rates were low. In fact, even though Fertig's intention was to be caught, he did not surrender voluntarily to the police. Quite

the contrary, Fertig was arrested because the criminal justice system's authorities were efficient in collecting the evidence that drove them to him. What should have happened, then, for Fertig's actions not to have taken place as they did? To put it bluntly, the complex social tissue in which Fertig's motivations were nurtured and engendered must have been radically different. For instance, the inexistent human commitment showed by the Mercy Memorial Hospital's workers to treat Fertig's son disease must have had to be present at least at a minimum rate. In case none of these social phenomena had actually bred Fertig's motivations, it is reasonable to presume that his owncriminal actions would not have been what they were.

As we know, this is a fictitious story in which the main character's motivations may seem at least understandable, even if not morally (or legally) justifiable. But consider, in contrast, the hundreds of gender-based violent crimes that sadly take place every day in our societies. What we tend to see as morally objectionable in those cases is not just the violence that ends up in taking an innocent human life or putting it at serious risk. It is also the outrageous motivations that underlie them, closely related to a millenarian culture that unfairly put women in an inferior scale of human evolution. If we believe that a criminal justice system can prevent by itself these kinds of crimes, for example by reinforcing the mechanisms of control or by intensifying the severity of the penalties, but acting as if the discriminatory elements incorporated in our culture over the course of the centuries were irrelevant, I seriously fear that our prospects are hopeless.

Changing the social tissue, as I said, can hardly ever be the saliently definitive function of any criminal justice system, which is just one system operating within a much bigger network integrated by other complex systems, legal and extra-legal as well (*cf.*Luhmann 1995, Chap. 10; Braithwaite and Pettit 1990: 82). At most, it can represent one of its possible functions, andespecially if it is defined in such a modest way as not to arouse false expectations. The alternative consensual model of a criminal justice system here proposed has the merit of allowing us to understand *why* to be modest and *to what extent*, for it signals that crimes that are the expression of an individual's willingness to be punished will usually have to be interpreted as the representation of human characteristics whose presence or absence could always lie far beyond the scope of the preventionist target. Against this background, our accountreminds us that the complex world of human motivations and the preventionist target of a criminal justice system do not necessarily go hand in hand. Precisely, the consensual ideal of a system of punishment operating *as a whole* sets the maximum scope of the *preventionist target*.[19] And, as such, it gives a first clue to begin to imagine how to act from within other social systems to prevent what the criminal justice system cannot prevent by itself.

V. Conclusions

In human affairs, as I said in the introductory part of this paper, different concepts of consent will be able to play different roles in different contexts, depending on what we purport to justify and how we pretend to do it. In political theory, for instance, a hypothetical notion of consent actually managed to do a magnificent job that probably would not have done in other areas. In legal theory, for its part, the notion that is typically at stake at the moment of justifying a given arrangement in tort law or in contract law – let us suppose – is doubtlessly more real and robust than the hypothetical notion, even if the required features it must reunite to acquire normative power will have to be evaluated on the

[19] I borrowed the expressions in italics from Braithwaite and Pettit 1990. For the preventionist target, the authors basically understand 'crime-prevention', which is "usually broken down into one or more sub-goals: the incapacitation or rehabilitation of actual offenders, for example, or the deterrence of potential offenders" (1990: 45). On the other hand, a criminal justice system operating as a whole must be read as including the operations that take place within each of the sub-systems that are part of it, such as the surveillance sub-system, the prosecutorial sub-system, or the sentencing sub-system, to mention some of the most important ones (*cf.* 1990: 7-8).

merits of each case. A tacit notion of consent will suffice in some contexts, whereas an explicit notion will be required in other, more demanding ones.

As regarding the criminal area of law, in this paper I tried to make it plain that, in principle, any notion will do, depending again on what we purport to justify. An intense willingness to be punished does not look odd at all when interpreted following Rawls'ss scheme, valid for what I called a **TSP**. Finkelstein's hypothetical notion of consent, for example, is very usefulat the normative design's political stage of a criminal justice system, but practically inert at its application stage. In contrast to these approaches, the notion of consent to punishment that I tried to rescue in the end of this paper is able to represent the individual's real will even in the context of a **NSP**. The role that it is supposed to play, however, is far from being a justificatory one, as the traditional consensual theories would have thought. In my view, it is better to conceive it as a purely expressive one. Generally speaking, when consent to punishment is predicated from the actions of an agent, what such a subjective attitude will express is that, in relation to certain types of offenses and in relation to certain ways of matching crimes with penalties, the preventionist target of a criminal justice system operating as a whole will see insurmountable limitations. Nonetheless, how that message can be unambiguously read in an individual's criminal actions will depend, among other things, on the level of law enforcement of that system. In conditions of less than perfect enforcement, as Randy Barnett once put it, the meaning of human actions will tend to be irreducibly ambiguous (*cf.* 1992: 902).

(A Kantian) Post-Script. The solution attempted in this paper as a way out of the Kantian dilemma has proved toexhibit no real influence from what might be deemed asa Kantian spirit. In a sense, that is a predictable outcome, since it was no other but Kant the one who addressed from the beginning the impossibility of there being someone willing to be punished. Despite this explicit refusal, I am confident that an authentic Kantian solution is still waiting to be brought to light from his writings. In the next conclusive remarks I would like to suggest, paraphrasing Habermas, how '*mit Kant gegen Kant denken*' canbe, after all, a promising enterprise.

In the *Groundwork*, Kant traced a distinction among three kinds of imperatives, namely: rules of skill, counsels of prudence, and commands (laws) of morality, which are also called 'technical (belonging to art)', 'pragmatic (belonging to welfare)' and 'moral (belonging to free conduct as such, that is, to morals)' (GMS 4: 416), respectively. With respect to the moral imperative, what Kant says is almost common knowledge among practical philosophers: it is categorical, for according to it "action is represented as objectively necessary of itself, without reference to another end" (GMS 4: 414). Regarding the other two, Kant says that they are hypothetical, for they "represent the practical necessity of a possible action as a means to achieving something else that one wills" (*ibid.*). In a previous section of this paper we had the opportunity to bring aboutthe principle that governs human conduct when rules of skill are at stake (see *supra*, 5). Now special attention must be givento the way Kant conceives the second kind of imperative. To his purpose, it is the contrast between it and the moral imperative that becomes salient. Hence, whereas "the categorical imperative (…) is limited by no condition and, as absolutely although practically necessary, can be called quite strictly a command"; on the contrary, "the necessity involved by counsels of prudence can hold only under a subjective and contingent condition, whether this or that man count this or that in his happiness" (GMS 4: 416). In the end, it is precisely the empirical and conditional character of this condition the responsible of ensuring the practical possibility of any pragmatic imperative.

If we are guided by this insightful reconstruction of our polyvalent normative agency – let us say, we easily get to the conclusion that agents such as the onesembodied in the characters of Fertig, Breivik, the Rawlsian civil disobedient, and others, are not inconceivable at all. Because according to Kant "happiness cannot be an ideal of reason but of imagination, resting merely upon empirical grounds" (GMS 4: 418), or subjective, or contingent ones (*cf.*GMS 4: 416), there seems to be no real impediment for him to envisage an agent who, in the name of her idiosyncratic well-being,

decides to rest on her own punishment as a means to further her end, even under a **NSP**. Of course, Kant proceeds with caution when it comes to recognize a *given* ideal of happiness (*cf.* GMS 4: 419). Such a thing is *futile*, Kant writes, once we realize that even the *mostinsightful* and *powerful* human being will face serious limitations to figure out a comprehensive and detailed picture of "what he really wills here" (GMS 4: 418), which would include, in his own terms, "the totality of a series of results in fact infinite" (GMS 4: 419). For the moment, however, this is irrelevant. What rather matters most is that "if it is supposed that the means to happiness can be assigned with certainty", therefore "the imperative of prudence would […] be an analytic practical proposition" (ibid.), recommending a course of action over whose desirability no one can pretend to have the final word. It has been my contention all along this paper that such a course of action, when it represents an offense or a crime within a given legal system and under certain conditions of enforcement, can express an authentic consent to punishment on the part of its perpetrator. Over and above Kant's explicit position thereon, I suspect that he would agree with it.

Bibliography

BEISER, Frederick: "Wolff, Chladenius, Meier: Enlightenment and hermeneutics", en J. MALPAS/ H.-H. GANDER (eds.), *The Routledge Companion to Hermeneutics*, New York, Routledge, 2015.

AMERSON, Richard: "Rawls versus Utilitarianism in the Light of Political Liberalism", in Victoria Davion and Clark Wolf (eds.), *The Idea of a Political Liberalism: Essays on Rawls*, Lanham, Rowman and Littlefield, 2000.

BARNETT, Randy E.: "The Sound of Silence: Default Rules and Contractual Consent", *Virginia Law Review*78 (1992) 821-911.

BARRY, Brian: *Justice as Impartiality*, Oxford, Clarendon Press, 1995.

BOONIN, David: *The Problem of Punishment*, Cambridge, Cambridge University Press, 2008.

BRAITHWAITE, John and Pettit, Philip: *Not Just Deserts: A Republican Theory of Criminal Justice*, Oxford, Clarendon Press, 1990.

BRAITHWAITE, John: *Restorative Justice and Responsive Regulation*, Oxford, Oxford University Press, 2002.

COHEN, G. A.: "Reason, Humanity, and the Moral Law", in Christine Korsgaard *et al.* (authors) *The Sources of Normativity*, Cambridge, Cambridge University Press, 1996.

DUFF, Anthony: *Punishment, Communication and Community,* New York, Oxford University Press, 2003.

ENGEN, Andy: "Communication, Expression and the Justification of Punishment", *Athens Journal of Humanities and Arts* (Oct., 2014) 299-307.

FEINBERG, Joel: "The Expressive Function of Punishment", in Joel Feinberg, *Doing and Deserving: Essays in the Theory of Responsibility*, New Jersey, Princeton University Press, 1971.

FINKELSTEIN, Claire: "A Contractarian Approach to Punishment", in Martin P. Golding and William A. Edmunson, *The Blackwell Guide to Philosophy of Law and Legal Theory*, Oxford, Blackwell Publishing, 2005.

———: "Punishment as Contract", *Ohio State Journal of Criminal Law*, Vol. 8 (2010) 319-340.

FOOT, Philippa: *Natural Goodness*, Oxford, Oxford University Press, 2001. Spanish translation by Ramón VilàVernis: *La bondad natural. Una visión naturalista de la ética*, Barcelona, Paidós Contextos, 2002.

FRANKFURT, Harry: "Identification and Wholeheartedness", in H. Frankfurt (author) *The Importance of What We Care About*, Cambridge, Cambridge University Press, 1998.

GARZÓN VALDÉS, Ernesto: "Consenso, racionalidad y legitimidad", *Isegoría* 2 (1990) 13-28.

GEUSS, Raymond: "Morality and Identity", in Christine Korsgaard*et al.* (authors) *The Sources of Normativity*, Cambridge, Cambridge University Press, 1996.

GOLDBERG, John C. P. and Zipursky, Benjamin C.: "Tort Law and Responsibility", in Goldberg *et al.* (editors), *Tort Law: Responsibilities and Redress*, New York, Wolters Kluwer Law & Business, 2012.

HART, H. L. A.: "Bentham on Legal Rights", in *Oxford Essays in Jurisprudence*, Alfred W. B. Simpson (editor), Oxford, Clarendon Press, 1973.

HILL, Thomas E., Jr.: "Kant on Punishment: A Coherence Mix of Deterrence and Retribution", in *Respect, Pluralism, and Justice: Kantian Perspectives*, Oxford, Oxford University Press, 2000.

HÖFFE, Otfried: *Politische Gerechtigkeit: Grundproblem der praktischen Philosophie*, Fernuniversität, Gesamthochschule, FachbereichErziehungs-, Sozial- u. Geisteswiss, 1989.

HONDERICH, Ted: *Punishment: The Supposed Justifications Revisited* (Revised edition), London, Pluto Press, 2006.

IMBRISEVIC, Miroslav: "The Consent Solution to Punishment and the Explicit Denial Objection", *Theoria*, 68 (2010) 211-224.

———: "Carlos Nino's Conception of Consent in Crime", *Diacrítica* vol.27 no.2 (2013)103-123.

KAHAN, Dan M.: "Social Influence, Social Meaning, and Deterrence", *Virgina Law Review* Vol. 83, Nº 2 (Mar. 1997) 249-395.

———: "Two Liberal Fallacies in the Hate Crimes Debate", *Law and Philosophy* 20 (2001) 175-193.

KANT, Immanuel: *Die Metaphysik der Sitten* (MS). English translation and edition by Mary Gregor: *The Metaphysics of Morals*, Cambridge, Cambridge University Press, 1996.

———: *GrundlegungzurMetaphysik der Sitten* (GMS). English translation and edition by Mary Gregor: *Groundwork of the Metaphysics of Morals*, Cambridge, Cambridge University Press, 1998.

KLEINIG, John: "The Nature of Consent", in Miller, F. & Wertheimer, A. (Eds.), *The Ethics of Consent*, Oxford, Oxford University Press, 2010.

KORSGAARD, Christine M.: "The Right to Lie: Kant on Dealing with Evil", *Philosophy and Public Affairs*, Vol. 15, Nº 4 (1986) 325-349.

KOTLER, Philip: Marketing Management, New Jersey, Prentice Hall, 1967. Spanish translation by Ramón Palazón Bertrán: *Dirección de mercadotecnia: análisis, planeación y control*, México D. F., Editorial Diana, 1973.

LLYOD THOMAS, D. A.: *Locke on Government*, London, Routledge, 1995.

LUHMANN, Niklas: *Das Recht der Gesellschaft*, Berlin, Suhrkamp Verlag, 1995.

MALAMUD GOTI, Jaime: *Suerte, moralidad y responsabilidad penal*, Buenos Aires, Hammurabi, 2008.

MORRIS, Herbert: "Persons and Punishment", in *Readings in the Philosophy of Law*, New Jersey: Prentice Hall, 2001.

MORRIS, Allison: "Critiquing the Critics: A Brief Response to Critics of Restorative Justice", *The British Journal of Criminology: An International Review of Crime and Society* 42 (2002) 596–615.

NINO, Carlos S: *Towards a General Strategy for Criminal Law Adjudication*, Oxford, unpublished PhD thesis, 1977 (for the revised Spanish translation, see Nino 1980).

———: *Los límites de la responsabilidad penal*, Buenos Aires, Astrea, 1980.

———: "The Consensual Theory of Punishment", *Philosophy and Public Affairs*, 12 (4) (1983) 289-306.

———: *The Ethics of Human Rights*, Oxford, Clarendon Press, 1991.

PEREDA, Carlos: "Lógica del consentimiento", in León Olivé (comp.), *Ética y diversidad cultural*, México, Fondo de Cultura Económica, 2004.

RAWLS, John: "Two Concepts of Rules", *The Philosophical Review* Vol. 64, Nº 1 (1955) 3-32.

———: "The Justification of Civil Disobedience", in Hugo Adam Bedau (ed.), *Civil Disobedience: Theory and Practice*, New York, Pegasus Books, 1969.

———*: A Theory of Justice*, Cambridge, Massachusetts, Harvard University Press, 1971.

RAZ, Joseph: *The Morality of Freedom*, Oxford, Clarendon Press, 1979.

———: "Government by Consent", in Joseph Raz, *Ethics in the Public Domain: Essays in the Morality of Law and Politics*, Oxford, Clarendon Press, 1987.

SCANLON, Thomas: "Contractualism and Utilitarianism", in AmartyaSen and Bernard Williams (editors), *Utilitarianism and Beyond*, New York, Cambridge University Press, 1982.

———: *What We Owe to Each Other*, Cambridge, Massachusetts: Harvard University Press, 1998.

SEN, Amartya: "Utilitarianism and Welfarism", *Journal of Philosophy* Vol. 76, nº 9 (1979) 463-489.

———: *Choice, Welfare and Measurement*, Cambridge, Basil Blackwell, 1982.

SHER, George: "Blame for Traits", *Noûs* 35:1 (2001) 146-161.

SIMMONS, John: *On the Age of Anarchy: Locke, Consent and the Limits of Society*, New Jersey, Princeton University Press, 1993.

URBAN WALKER, Margaret: *Moral Repair: Reconstructing Moral Relations after Wrongdoing*, Cambridge, Cambridge University Press, 2006.

WALL, Steven: *Liberalism, Perfectionism and Restraint*, Cambridge, Cambridge University Press, 1998.

WENNBERG, Mikko: "Modeling Hypothetical Consent", *Journal of Libertarian Studies* Vol. 17, Nº 3 (2003) 17-34.

WILLIAMS, Bernard: "A Critique of Utilitarianism", in J. J. C. Smart and B. Williams, *Utilitarianism: For and Against*, New York, Cambridge University Press, 1973.

———: *Shame and Necessity*, Berkeley, University of California Press, 1993.

———: "Internal Reasons and the Obscurity of Blame", in *Making Sense of Humanity and Other Philosophical Papers*, Cambridge, Cambridge University Press, 1995.

YURICK, Sol: *Fertig*, New York, Trident Press, 1966.

ZAFFARONI, Eugenio Raúl: "¿Vale la pena?", in *Controversias de derecho penal*, Bahía Blanca:Programma Nº 1 (2005) 41-59.