Volume 2   ISSN: 2585-3899
Issue 1
2019
Special issue:
Blockchain and disruptive technologies:
Interdisciplinary perspectives
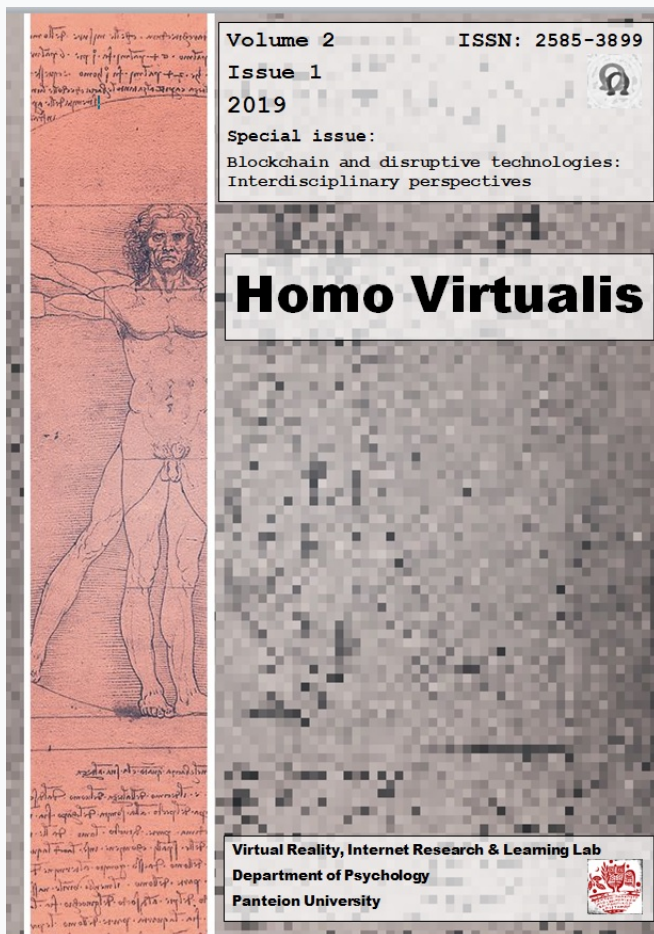
Homo Virtualis

Virtual Reality, Internet Research & Learning Lab
Department of Psychology
Panteion University

Surveillance, social punishment, and the viability of Homo Virtualis

Robert C. Mellon

To cite this article:
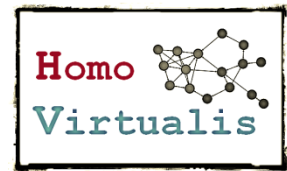
Surveillance, social punishment, and the viability of *Homo Virtualis*

Robert C. Mellon[1]

Abstract: As emerging encryption technology results in widespread freedom from detection of online activity, *Homo Virtualis* might usefully be considered as *Homo sapiens* without a gun to its head—a highly evolved species experiencing, for the first time on a massive scale, a home environment in which the individual who engages in acts that harm others is not identifiable. This renders the punishment of antisocial behavior in this context impossible, clearing the warning signals for retribution and crippling this ancient and brutal method for limiting the occurrence of antisocial or unwanted behavior. If *Homo Virtualis* is to retain this newfound freedom from threat and invigilation, prosocial behavior must continue to prevail in the absence of credible threat of punishment for the antisocial. The prospects are far from certain, and an attendant spike in the frequency of misdeed would be sure to evoke a crackdown on encryption tech. On the bright side, a widespread failure of punishment to inhibit malice and malfeasance online might finally evoke serious consideration of how *H. sapiens* might ensure a low frequency of antisocial behavior absent the threat of dire consequence for transgression, via the positive reinforcement of prosocial conduct, rendering interpersonal and community sensitivity and service pleasurable rather than obligatory. Encryption technologies can provide valuable support for such a comprehensively humanistic effort to reduce malfeasance by eliminating threats of public censure for current and budding offenders' own efforts to address their problematic proclivities.

Keywords: *social control of behavior, social punishment, encryption technology, surveillance technology, behavioral engineering*

Homo sapien's tradition of aversive social control of undesirable behavior

In common with the members of many other species, throughout recorded history human beings have arranged a regimen of threat, surveillance, and inflicted pain of reduced circumstances as a means of reducing the probability that their conspecifics will act in ways that harm themselves and others and conflict with the interests of the community. This is the model of social control of individual conduct that we humans employ when we threaten, say, confinement for a given period as a penalty for engaging in a certain activity, then

---

[1] Professor and Director of Laboratory of Experimental and Applied Behavior Analysis, Dept. of Psychology, Panteion University of Social and Political Sciences, Athens, Greece. E-mail: mellon@panteion.gr; mellon.robert@gmail.gr

collect observations that, if judged to reliably indicate that the person is or has behaved in this manner, compel us to restrict her or his wellbeing in accordance with prevailing law or community practice.

This ancient and brutal method of reducing the probability of occurrence of undesirable behavior in others is commonly employed in our savage markets and labor relations, in our cutthroat educational systems, in apocalyptic religious and ethical training and in our everyday disagreements with neighbors, friends and family members. Even infatuated lovers may be found to censure their sweethearts' inattention by petulant threats of withdrawal of affection. The principal difference between the method of control of undesirable behavior in our explicitly codified legal systems and in our tolerably menacing interpersonal control practices would appear to be the intensity of displeasure experienced by both the aggrieved and the reproved parties; indeed, it is typically when interpersonal threat and censure fail to control undesirable behavior in others that we turn to a legal system publically authorized to intervene with more powerful threats of aversive consequences for unsanctioned activity.

Limiting social misconduct by the tripartite method of threat, surveillance and punishment of noncompliance is costly even for those who scrupulously obey the rules, for the innocent must also submit to public scrutiny of activity and suffer the same perpetual threat of dire consequences for transgressions that have not yet committed—but still might. Nobody wants to be cautioned, monitored, nor least of all transitioned to a lesser state of wellbeing, but in order to limit the malfeasance and cruelty of others, antediluvian social convention demands that we each in turn expose ourselves to public inspection and possible censure for actions observed or implied.

If the threat is sufficiently potent, aversive social control often does produce high levels of self-vigilance for the warning signals from the inchoate occurrence of unethical, inept, inapt, or otherwise punishable actions and thoughts, resulting in precise if inflexible adherence to explicit or implied rules of acceptable social and intrapersonal conduct. The universally popular use of social punishment is based in the commonplace effectiveness of making others afraid or miserable until they do what they are told. But the control of behavior by contingent relief from threat of transition to a lesser state is distressing for both parties, even when it works as planned. And when effective, high levels of watchfulness for potential errors and unintended desecrations in one's own thoughts and deeds can readily displace essential activities such as experimentation and rational risk-taking; behavioral conditions such as obsessive-compulsive disorder, panic disorder and paranoid ideation may be manifestations of generalized vigilance for potent, if subtle, warning signals produced by one's own censurable thoughts and actions (e.g., Mellon, 2009; 2013).

When the threat of punishment fails to reduce the probability of offensive behavior, as it often does, everyone pays a penalty. Conformity or compliance is never the only way to terminate unpleasant warnings or demands (e.g., Skinner, 1971; Sidman, 1989) and many alternative forms of escape from threat prove to be more problematic than the original transgressions. Most obvious among them are the myriad forms of counter-threat and counter-attack, activities that expose both parties to elevated risk of harm or displeasure; but sullen contempt, pitiable displays of inadequacy or victimhood, anarchy, nihilism, self-harm or attempted suicide, abuse of drink, psychoactive drugs, and food—and lately, withdrawal into cyberspace—are also among the common forms of pernicious escape occasioned by everyday aversive social control.

Aversive social control of behavior is derailed by anonymity

Yet another unloved offspring of oppressive behavior control practices is deceptive or cryptic misconduct. Threats of punishment for undesirable activity can be terminated by acts that render the activity itself unobservable. Lying about personal responsibility, for example, is a form of escape from threat of social punishment that is effective only when the identity of the author of an offence cannot be detected by the aggrieved party. In a system of control based upon threat of transition to reduced circumstances, anonymity is tantamount to immunity from social punishment, a precious commodity created by aversive control practices themselves and, ironically, proving itself to be a source evolution of antisocial activity. For example, in the hope of reducing street crime, a community might sacrifice their own privacy and install security cameras, but the well-prepared lawbreaker can hack the system by wearing anti-surveillance clothing, which in turn has led police officers to selectively interrogate young men who wear hoodies, and even penalize their public donning, sparking protests of civil rights violations (e.g., Patterson, 2016)—and so on.

The effects of even partial social camouflage on the probability of occurrence of punishable behavior is clearly visible in the comments section of a platform like Youtube, where commenters are identified only by a username of their own choosing. In this environment, evaluations about performances, opinions, group membership, physical appearance and the like are far more nasty and cruel than those typically encountered in eponymous interactions, in which such unkind behavior tends to have powerful repercussions. Surveys of users of such platforms have indicated that as many as one in four admit to have engaged in intensely aversive commentary (e.g., March & Marrington, 2019). With increasing frequency of use, words and phrases that once elicited shock and evoked attention become ineffective or even boring, such that the typed productions of persons producing "likes" for "restoring justice" by hurting themselves and others inevitably become more and more impressively hurtful. We are each unkind in accordance with the history of effects of our unkind acts in a given context, and anonymity is a context in which many long-standing sanctions against incivility are lifted.

Of course not all usernames are equally anonymous or plausibly deniable, and no one who is impressed by the low-tech hacking of a Guccifer could believe in truly incognito internet trolling or bullying, but it is all the more disheartening to see what has happened to public discourse by simply rendering the discovery of the authors of verbal assault *inconvenient*. It is disheartening in its suggestion that in contexts in which rational dialogue of civil tone *is* observed, it is not sustained so much by its positive virtues as by its role in averting threats of retaliation for rudeness. That suggestion well and truly shakes our confidence in the viability of *Homo Virtualis* in the epoch of advanced digital encryptology now upon us.

The struggle for freedom and the portent of a darker web

As information and economic scientists writing in this issue and in the global forum attest, efforts to generate a convenient, permanent, impenetrable, incorruptible and sustainable digital monetary systems have given rise to distributed encryption technologies that can be utilized to mightily challenge the identification of parties engaged in monetary transactions and other forms of social exchange that are conventionally met with censure (e.g., Whyte,

2019). These include not only crimes such as money laundering, child abuse, human trafficking and the funding of terrorist activities (e.g., Foley et al., 2018; Whyte, 2019), but also legal activities that are met with disapproval and derision when observed by many people, such as groups encouraging hate speech, self-harm and nihilistic or suicidal ideation in young people (e.g., Daine et al., 2013).

In his analysis of the history and determinants of the social struggle for freedom, B. F. Skinner (1971) noted that this enduring fight of and for humanity specifically concerns the termination of aversive control—of forced choice, compelled conduct, and the policing of self that is requisite to escape the threat of others. Nobody protests when their actions occasion accommodation, respect, mutual benefit and pleasure from others, though such consequences of our acts obviously tend to increase the probability of occurrence of similar activity on subsequent occasions; operant behavior just works that way. No, in our personal as in our collective struggles, we specifically yearn for freedom from unnecessary and painful or unwanted social imposition. With respect to our online activity, that is the condition currently promised by emerging encryption technology. It comes with a catch.

To remain free, *Homo Virtualis* must continue to behave with humanity in the absence of any threat of detection—in the perfect conditions for extracting brutish satisfaction at other peoples' expense. Otherwise, the threat of punishment will not only return, it will likely be redoubled. Decentralized cryptocurrencies have already been banned in a number of countries, principally in states known to restrict and surveil their citizens' activities on the internet. That can happen anywhere if adequately provoked. It may be accompanied by exposure to propaganda in which the viewer is initially charmed by the protagonist, who falls on very hard times when lured into encryption violation. Thus the watchful citizen is taught to recognize and terminate his or her own exposure to environments that are temptingly free of the threat of punishment. The *virtualis* species of hominid will thus either fall victim to the intrapersonal oppression of its own conduct, reduced to a pallid encryptophobe, or it will go rogue in a struggle to restore freedom in cyberspace, which is no picnic either.

Should the termination of the threat of punishment for online activity fail to occasion behavior characterized by enlightened self-interest—should we fail the test of newfound freedom—the size and scope of the ensuing crackdown on cryptology will depend upon the new technology's sustained resistance to cracking. Many in the cyber community have high hopes about the refuge potential of blockchain technology, and if it is not the promised cloak of invisibility, superior blurring of identity may suffice to occasion dramatic change in online activity. With surveillance derailed and misdeed in the ascendant, the only way for society to maintain the effectiveness a debilitated aversive control is to increase the painfulness of punishment and the potency of its threat (read: "Winter is coming").

One cannot but think that, at present, the viability of *Homo Virtualis* as *Homo sapiens* truly free from social sanctions is suspect. There is, however, hope that a truly private internet will serve as a sort of laboratory for how things might be arranged such that human beings tend to behave in accord with the public good absent the threat of punishment for misdeed. Today and whenever cryptology renders pernicious behavior undetectable, as an alternative to more extreme sanctions, we might think how we might come to live well, online and off, without fear of retribution for error or misconduct. A natural science of behavior can help with that (e.g., Skinner, 1953; Sidman, 1989; Μέλλον, 2013).

How a darker web might expedite behavioral enlightenment

How do we balance the individual's right to privacy and freedom from punishment with the necessity of limiting the occurrence of malice and malfeasance? The *positive* reinforcement of behavior incompatible with problematic activity is an attractively potent alternative. Over time, the probability of occurrence of any given form of antisocial behavior is reduced whenever the consequence of an incompatible prosocial behavior is transition to a better circumstance. That is true whether the reinforcing transition is the *termination* of a threat (*negative* reinforcement) or the *production* of a welcomed condition (*positive* reinforcement). This is the difference between *having* to do something and *wanting* to do it—between fear of displeasure if not performed vs. anticipation of pleasure when performed.

If we can no longer detect the activity of a person who behaves badly in a context of virtual anonymity, perhaps we will finally get around to a more effective intervention strategy than our bronze-age policy of catch, cudgel and cage. One of the splendid features of positively reinforcing prosocial behavior is that the images, words and other events produced by our own behavior when we solve problems rationally and justly acquire positive value or reinforcing potency, such that we enjoy doing the right thing for ourselves and our community even when no one is watching. We need no intimidating surveillance to make us keep doing what we want to do—the trick is to establish the positive reinforcing potency of the events produced by behaviors that serve the public good, such as by celebrating the correspondence of explanation and evidence, and the consideration of the effects of our behavior on those around us, both immediately and with the pass of time. These and other unequivocally prosocial characteristics of behavior are useful to others, who may well show their appreciation, further strengthening and refining them.

By establishing the positive reinforcing potency of doing the right thing—and of systematically examining why we judge that thing it to be right—we might protect the common good without menace and supervision. It may be too late to reform the unsupervised misbehavior of some of us who never learned to enjoy and value the welfare of those around us—and every bad action will increase the probability of a general crackdown on private interaction—but most us who have never before lived in freedom from punishment are still governed by pain of conscious trained in extended aversive surveillance, and we tend to distrust promises of anonymity too much to risk malfeasance. Upcoming generations will know less of this and will be increasingly unable to depend upon social censure and its threat to curb their malicious tendencies. To thrive in this brave new world of anonymity they must well and truly come to behave beauteously even when not forced to do so, because in cyberspace at least, they might never be forced to do so. Otherwise, we can only expect the worst from a darker corner of cyberspace where, for the moment, our offending sister and brother humans choose to exercise the dubious freedom afforded to a fugitive from social punishment.

True anonymity also reduces threats posed by positive prosocial intervention

To recap, positively reinforcing prosocial behavior reduces the frequency of occurrence of antisocial behavior without threat of pain and anxious vigilance. A derailing of the aversive control of behavior by advanced cryptography might thus compel us to rethink how we arrange circumstances to reduce the frequency of occurrence of problematic behavior. Beyond necessitating that we step beyond brutal social conventions to limit problematic activi-

ty rendered undetectable, a truly private electronic interaction platform could itself contribute significantly to the dissemination of positive reinforcement-based social interventions.

People who fantasize and engage in widely censured activities such as sexual harassment and assault, child abuse, self-abuse, and hatefulness to groups are naturally reticent to disclose their behavior to anyone not known to share their proclivities. As a result, they speak of their problematic behavior principally to persons who are likely to encourage it. The acceptance of widely hated personal qualities that social offenders find in marginalized and aggrieved group interaction readily generates loyalty to the clandestine clan, further reducing the probability that the individual will betray their problematic behavior to someone who can help to change its course. But the ready availability of a truly anonymous channel of human interaction will insulate the exploration and discussion of the nature and determinants of pernicious activity from the censure of both the in-group and the general population, removing a fundamental barrier to the efforts of specialists in problematic behavior to help the person to get their life back on track.

In our primitive blood lust to avenge wrongdoing, we often fail to recognize that the antisocial behavior that damages, disappoints or revolts us is deeply rooted in the individual's history of interaction with the world, and that the future course of a harmful action tendency depends also upon how we treat the person's behavior now. In bitter irony, the cruel and hurtful behavior that we are so anxious to punish in others is often itself a product of aversive control. Helping a person to understand how, in the course of their life, the effects of activities that harm others acquired positive value can reduce their reinforcing potency.

Likewise, helping a person to gain what is needed from others in a manner that is mutually advantageous establishes the positive value of events produced by socially responsible behavior—by acts physically incompatible with misdeed. Helping, say, a pederast to understand how such dreadful behavior is shaped in events of living, and to discover the irresistible joy of mutually fulfilling sexual activity in the context of an intimate relation with an intellectual and experiential peer, cannot guarantee the eradication of his malfeasance, but it remains our best chance—and perhaps soon our only chance—to produce the kind of behavior that characterizes enlightened self-interest in a social space radically freed of hostile imposition. The very real threat of encryption technology might thus be transformed into an opportunity for the erosion of threat of punishment itself as our go-to policy for limiting malfeasance. *Bonus fortuna, Homo Virtualis!*

References

Daine, K., Hawton, K., Singaravelu, V., Stewart, A., Simkin, S., Montgomery, P. (2013). The power of the web: A systematic review of studies of the influence of the internet on self-harm and suicide in young people. *PLoS ONE*, *8,* e77555.

Foley, S., Karlsen, J. R., and Putnins, T. J. (in press). Sex, drugs, and Bitcoin: How much illegal activity is financed through Cryptocurrencies? *Review of Financial Studies.* Retrieved from http://dx.doi.org/10.2139/ssrn.3102645

March, E., & Marrington, J. (2019). A qualitative analysis of internet trolling. *Cyberpsychology, behavior and social networking, 22*, 192-197.

Mellon, R. C. (2009). Superstitious perception: Adventitious reinforcement and punishment as determinants of repetitive eccentric interpretations. *Behaviour Research and Therapy, 47*, 868-875.

Mellon, R. C. (2013). Interpreting thought-action fusion in obsessive thinking: Covert links of operant chains in clinical behavior analysis. *European Journal of Behavior Analysis, 14*, 177-197.

Moeser, M. (2013) Anonymity of bitcoin transactions: An analysis of mixing services. Proceedings of the Muenster Bitcoin Conference. https://www.wi.unimuenster.de/sites/wi/files/public/department/itsecurity/mbc13/mbc13-moeser-paper.pdf

Sidman, M. (1989). *Coercion and its fallout*. Boston: Authors Cooperative.

Skinner, B. F. (1953). *Science and human behavior.* New York: Macmillan.

Skinner, B. F. (1971). *Beyond freedom and dignity*. New York: Knopf.

Whyte, C. (2019) Cryptoterrorism: Assessing the utility of blockchain technologies for terrorist enterprise. *Studies in Conflict & Terrorism,* DOI: 10.1080/1057610X.2018.1531565

Μέλλον, Ρ. (2013). *Ψυχολογία της συμπεριφοράς*. Αθήνα: Πεδίο.

Notes on Contributor

Robert C. Mellon, Ph.D. is Professor in the Department of Psychology at Panteion University in Athens, where he directs the Laboratory of Experimental and Applied Behavior Analysis. After undergraduate studies in psychology at the University of North Carolina at Charlotte, undertook master's and doctoral training at the University of North Carolina at Greensboro, where he completed the requirements for both the Doctoral in Clinical Psychology and Experimental Psychology programs. He was postdoctoral fellow at the Center for Developmental Psychobiology in the Department of Psychology at the State University of New York at Binghamton, and was National Institute of Mental Health Research Service Award Fellow in the Department of Biopsychology, N.Y. State Psychiatric Institute and Columbia University. He has served on the faculties of the University of Maryland and the University of Crete and as president of the European Association for Behavior Analysis and the Hellenic Association for Behavior Analysis.