Integrity, Shame and Self-Rationalization

Elias L. Khalil

Vassar College Economics Working Paper # 55

ABSTRACT

In orthodox economics, honesty or integrity is treated as either part of the constraint function or the objective function. This is at the origin of the statement, “every man has his price.” However, if integrity has a price, why do agents experience shame when they sell it while they do not when they sell other possessions? If agents are rational, why do they resort to self-rationalization to avoid shame? The proposed view, called “quantum,” avoids these anomalies. Further, the quantum view avoids another set of anomalies that face heterodox models based on the multiple-self framework. The quantum view sheds light on heroism, self-identity, self-cheating (procrastination), etiquettes, taboos, and identity switch.

Key words: Cretan Liar Paradox, identity, identity switch, heroism, imprinting and the transactional view (pragmatism), self-cheating (procrastination), fairness, unitary-self view, multiple-self view, quantum view

JEL classification: D0
<table>
<thead>
<tr>
<th>Is Public Exposure Necessary?</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Is Action YES</strong></td>
<td>Shame of Ridicule: Indecent exposure that hinders “reflexive utility” related to decorum</td>
<td>Shame of Opportunism: Free-riding that buttresses “substantive utility” related to promises</td>
</tr>
<tr>
<td><strong>the Object NO</strong></td>
<td>Humiliation</td>
<td>Ignominy</td>
</tr>
<tr>
<td><strong>Of Evaluation?</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 1: Self-Disapproval in Backward-Looking Evaluations (Integrity Utility)*
Introduction

Economic theorists are generally pressed hard to account for the emotional and physiological state known as “shame”—despite its ubiquity. Sex offenders, embezzlers, collaborators with the enemy, shop lifters, and so on, try to cover their faces from a curious public. As Michael Lewis puts it, the person who feels shameful usually wants to hide, disappear, or even die:

Shame results when an individual judges his or her actions as a failure in regard to his or her standards, rules and goals … The person experiencing shame wishes to hide, disappear or die. It is highly negative and painful state that also disrupts ongoing behavior and causes confusion in thought and an inability to speak. The body of the shamed person seems to shrink, as if to disappear from the eye of the self or others. Because of the intensity of this emotional state, and the global attack on the self-system, all that individuals can do when presented with such a state is to attempt to rid themselves of it [Lewis, 1995a, p. 71; see also Lewis, 1995b, p. 75].

Shame usually arises from the violation, as Lewis puts it, of one’s own standards of moral behavior—standards aimed to buttress one’s own sense of integrity:

Definition: Shame is a disutility of a particular kind that arises from most, but not all violations of integrity.¹

¹ In this light, guilt can be seen as a milder form of shame [Lewis, 1995b, p. 75-77; Tangey, 1990, p. 103].
In turn, we need to define integrity:

**Definition:** Integrity is a utility of a particular kind that usually arises from backward-looking, intertemporal action that maintains the unity of the self. Integrity of the self is affirmed when the “present self” honors, when it has a choice to dishonor, a commitment made by a “past self.” The present act of honoring has no meaning without referring to the context, viz., the commitment or promise made by the past self when the present self can defect. As such, the context cannot be divided at the margin if intertemporal unity, i.e., integrity of the self is to be maintained.

The exact ramifications of these definitions are the subject of this essay. It is important to note here that these definitions do not allude to any particular cultural content. Integrity assures the preservation of identity, which can be considered a capital stock. George Akerlof and Rachel Kranton [2000] model identity as a taste. However, identity *qua* integrity is a capital stock, called below “identity capital,” which is necessary for the constitution of the constraint function irrespective of the historical or cultural setting. Such a setting only defines the exact details of integrity *qua* identity capital.² Although the essay focuses mainly on integrity *qua*

² What defines honesty, integrity, honor, and shame vary across individuals and across cultures [Cohen & Nisbett, 1994, 1997; Mosquera et al., 2002]. An act committed against a clan member might be considered dishonest, but if committed against outsiders, it might be considered “clever.” This difference is probably one aspect of ethnic and clan solidarity. This paper is not concerned with historical and cultural peculiarities. In fact, to come to grip with these
utility (of a special kind), the simple model offered here treats integrity as a constraint.

The preservation of identity, and the avoidance of shame, is so central to human and physical mental health that the failure may lead to multiple personality disorders as discussed by Lewis [1995a,b]. The preservation of identity is central enough that agents regularly resort to strategies of denial in order to suppress paradoxically from their own consciousness any hint of shame. One common strategy is self-rationalization [see Broucek, 1991; Schneider, 1992; Miller, 1996], which can be defined in the following way:

**Definition:** Self-rationalization amounts to the over-stretching of a narrative to convince mainly the self that the self did not sacrifice integrity for a higher income or greater comfort.

As the story of Adam and Eve illustrates, the agent usually tries to rationalize actions already taken in order to convince mainly himself that he is a man of integrity. The agent even may resort to “projection,” i.e., thinking that others are dishonest and even evil, as Hanah Arendt [1976] has argued with regard to the indictment of Nazi war criminals.

Social psychologists have amassed enormous empirical findings about shame and the pathologies that it can generate [e.g., Buss, 1980; Nathanson, 1992; Morrison, 1996; Lansky & Morrison, 1997; Gilbert & Andrews, 1998; Lewis & Haviland-Jones, 2000; Seidler, 2000; Tangney & Dearing, 2002]. Sociologists have studied the phenomenon of shame in relation to violence and social deviance [Scheff 1990; Retzinger, 1991; Scheff & Retzinger, 1991]. On the peculiarities, we need a simple model that captures the relation of integrity and interest in an ahistorical sense.
other hand, anthropologists have identified shame as an important cement of communal
solidarity [e.g., Gilmore, 1987; Hans, 1991]. Some political scientists have also highlighted how
shame gives origin to hubris and the quest after political power [Fisher, 1992]. Economists have
started to pay attention to shame, usually under terms such as “honesty,” “trust,” “fairness,”
“retaliation,” and so on, as exhibited in the ultimatum, trust, and dictator games [Fehr & Gächter
in Ben-Ner & Putterman, 1998; Levine, 1998; *passim* Khalil, 2003a].

While the approaches of the economists are more analytical than the discussions found in
other disciplines, the analytical rigor comes at the expense of richness of understanding the
psychological and cultural meaning of shame. This paper attempts to preserve the deep
understanding of shame gained from the other disciplines—but without sacrificing the analytical
rigor that typifies the economic approach.

In fact, this paper finds the theoretical treatment of honesty in economics to be less than
satisfactory for a number of reasons. This paper shows that the mainstream approach, called
here the “unitary-self” view, faces different anomalies. Mainly, the unitary-self view, with its
diverse flavors, can neither explain shame nor self-rationalization at first approximation. The
heterodox critics of the mainstream approach, grouped here under the “multiple-self” view, are
capable of explaining shame and self-rationalization. However, once the promoters of the
multiple-self view split the agent into multiple selves, they encounter difficulty in putting the
selves back together.

The paper proposes an alternative view, called “quantum,” which promises to supersede
the failings of the unitary- and multiple-self approaches. The quantum view proposes that while
self-integrity is a taste, it cannot, contrary to the unitary-self theories, be part of ordinary tastes that make up the usual utility function, what is called here “substantive utility.” Rather, integrity is the meaning that defines the context of acting according to one’s identity, i.e., affirming the continuity between present self and past self. As a meaning, integrity cannot be divided at the margin as ordinary tastes à la unitary-self theory. Of more importance, as a context, integrity cannot be an element of the ordinary utility set that affords context.

To elaborate, integrity as an element in the ordinary utility set would invite a logical contradiction: Integrity qua context has a zero cost of production. The only cost integrity has is the foregone gain from acting opportunistically. But such a cost is an opportunity cost of the non-ordinary kind. Namely, integrity is a by-product of a choice over alternatives taken within a context and, hence, integrity cannot be reduced and made into one of the alternatives. Such a reduction would make the set of alternatives a member of itself, which invokes the Cretan Liar Paradox [Russell, 1956].

On the other hand, the quantum view proposes that while self-integrity is non-ordinary, it cannot, contrary to the multiple-self theories, be totally separate from ordinary tastes that make up substantive utility. That is, integrity qua context entails that it cannot be totally divorced from its substrate—as much as the meaning of a sentence cannot exist independently of the words that make up the sentence. The term “quantum” is chosen exactly to distance the taste for integrity from these two kinds of modeling.

Section one reviews and criticizes standard (i.e., unitary-self) theories and heterodox (i.e., multiple-self) theories. Section two provides conceptual clarifications of the scope of shame.
Section three lays out the quantum model of shame. Section four reviews some ramifications of the quantum approach with regard to identity switch, imprinting and the transactional view of action (pragmatism), heroism, procrastination (self-cheating), the penal code, taboos, and etiquettes. The last section draws highlights the main point.

1. Theories of Integrity

1.1 Unitary-Self Approach

1.1.1 The integrity-as-strategy view

The standard, unitary-self view comes in three varieties [Khalil, 2003a]. The first one models integrity as a cost-effective constraint that the agent adopts to win the trust of others in repeated games [e.g., Tullock, 1985; Axelrod & Dion, 1988]. The games do not have to be infinite, but long enough to confirm that the present cost of defection outweighs the present cost of cooperation [Kreps et al., 1982]. In this, what one may call, “integrity-as-strategy” view, agents act with integrity because they are enlightened egoists.

The integrity-as-strategy view definitely explains why many firms forego myopic profits in order to buttress “goodwill.” But this view cannot explain at first approximation one anomaly: the persistence of integrity in one-shot games. At a second approximation, though, the anomaly can be explained. To decipher whether the game is one-shot or repeated involves sufficiently high transaction cost. So, the agent opts for a heuristic of acting with integrity in all cases because, on average, the foregone opportunistic gain is lower than the transaction cost needed in order to differentiate among games case-by-case.
This secondary approximation, i.e., the transaction cost qualification, may succeed in explaining why excessive calculation may convince one to behave always with integrity. Still, however, there are games, such as leaving tips in far-from-home restaurants or experimental games such as the ultimatum game, that clearly involve zero transaction cost. The agent obviously should defect and leave no tip in a far-from-home restaurant—which is generally not the case [Azar, 2004]. The agent obviously should take any proposed amount of money on the table in the ultimatum game—which is generally not the case [Güth et al., 1982; Güth et al., 1990; Bolle, 1995, 1996; Roth et al., 1991; Roth, 1995, pp. 258-260; Camerer & Thaler, 1995].

In particular, in the ultimatum game, the Proposor expects the Respondent to be despondent, i.e., act vengefully and refuse the proposed division of the money if the division is deemed “unfair.” The Proposor acts fairly probably because the Respondent is ready to hurt his substantive utility, i.e., is ready to act with vengeance [Posner, 2004]. ³ In either case, whether

³ Gary Bolton [1991] disputes the vengeance motive in the ultimatum game. He argues that the Responder's utility in the ultimatum game consists of absolute income as well as relative income, where low income relative to others engenders envy. He argues that a Responder rejects a low offer when the disutility of inequality outweighs the utility of the absolute amount offered. Such an explanation, however, is appropriate if the inequality is generated by a random device, not by a Proposer. To wit, experiments show that Responders tend to accept low offers when, first, generated by a computer [Rabin, 1993] and, second, generated as a result of competition among Responders [Roth et al., 1991]. Thus, Responders resort to punishment when they attribute the low offer to unfairness on the part of Proposers; they do
the Proposor wants to be fair or the Respondent wants to level punishment, the agent should not behave as such, given zero transaction costs, according to the integrity-as-strategy view.

1.1.2 The integrity-as-taste view

So, how to explain the phenomenon of integrity in single-shot games (with close to zero transaction cost)? To solve the anomaly, other theorists have placed integrity in the objective function, i.e., as an ordinary taste [e.g., Andreoni, 1988; Rabin, 1993; Levine, 1998]. According to what one may call the “integrity-as-taste” view, integrity can be sacrificed at the margin as the intensity of temptations rises. The integrity-as-taste view is epitomized in the cliché “every man has a price”—which, curiously, connotes a disdainful or a repugnant judgment of dishonesty.

As discussed shortly, this connotation is the Achilles’ heels of the integrity-as-taste view. But one should recognize that this view solves the anomaly facing the taste-as-strategy view; i.e., it can explain integrity in single-shot games. It solves it, though, by creating a host of other anomalies.

The main failing of the integrity-as-taste view is not that it violates the desired, standard assumption of stable preferences [Stigler & Becker, 1977]. The theorist can always claim that he has “forgotten” to include the taste for integrity in the original formulation of the utility function. The main failing rather consists of two anomalies, shame and viability.

It may seem odd to call shame an anomaly. It is obviously a disutility that arises when

---

not merely punish in order to avoid the disutility of envy.
one sells a particular commodity, integrity. Thus, as the integrity-as taste view maintains, shame is not different from any other disutility that arises when one sells, e.g., a table or a labor service. However, shame is not an ordinary disutility—as much as integrity is not an ordinary commodity that engenders an ordinary utility. If integrity is an ordinary utility or taste, the disutility arising from the sale of integrity should be announced and celebrated when the fetched price is high enough—similar to the announcement of any hard work or sacrifice when the fetched gain is high enough. The celebration of dishonesty is not suppressed because of fear of retaliation: We are discussing here the single-shot game where retaliation is not an issue.

In any case, if the reason why one does not celebrate the hard work of shame, as one does the hard work of labor, is the fear of retaliation, we have a prediction: The dishonest agent should at least confide to himself how proud he is for being a rational liar, i.e., for lying when expected opportunistic gain is greater than expected cost in terms of retaliation and disutility of shame. But stylized facts generally do not support this prediction: The dishonest agent may even resort to what psychologists call “projection” when he expresses the cliché “every man has a price” when he judges the motive of others. In projection, the agent subconsciously attributes his own dishonesty onto others in order, paradoxically, to hide from himself such a judgment [see Grant & Crawley, 2002]. So, why is the agent trying to hide from himself the fact that he is a rational liar? He should, at least in private, be proud of his efficiency.

So, the question persists, if every man has a price, why do people try to hide from others in single-shot game—where the fear of retaliation or the protection of reputation is not of concern? A more puzzling anomaly, why do people try to hide their opportunistic act from their
own selves when they stretch the facts of a narrative, i.e., what is known as self-rationalization? Even thieves and hardened criminals resort to self-rationalization—such as the victim “had it coming”—as they attempt to avoid cognitive dissonance [Akerlof & Dickens, 1982].

It seems that the agent’s shame cannot be offset or justified by the level of opportunistic gain—otherwise why undertake self-rationalization? Of course, in cases of emergency, i.e., in cases of life-and-death as shown below, dishonesty is justified. But beyond the scope of emergency, shame arising from acting contrary to a promise would be usually amplified further, rather than muffled, when one justifies it by the enormity of the opportunistic gain. The phenomenon of shame, therefore, is a thorn in the face of the integrity-as-taste approach.

There is a good reason why shame arising from the sale of integrity differs from ordinary disutility. Integrity is not an ordinary commodity such as A, B, and C. Integrity is rather the context of choosing A rather than choosing B or C. The context is the promise made by the past self. To treat the context of choosing A as lying along the same continuum, i.e., unitary utility, with A itself amounts to treating the set of choice as a member of itself. This invokes the logical contradiction mentioned earlier, i.e., the Cretan Liar Paradox.

Another strategy of self-rationalization is that all others seem to be cheating as well. This strategy, though, may not be self-rationalization if the charge is true. One is not obligated to be exploited as in the case of paying for a public good while all others free-ride [Margolis, 1991].

To note, the anomalies of shame and self-rationalization do not escape the model of Dowell, Goldfarb, and Griffith [1998]—despite their conscious effort to distance themselves from the
Concerning the second anomaly facing the unitary-self approach, i.e., viability, people with a taste for integrity are similar to people with a taste for pain or a taste for kerosene. Such tastes make the agent a non-viable entity, prone to sickness and, hence, cannot survive in market competition. In market competition where single-shot games are paramount, the sincere may inherit the heavens, but he cannot inherit the earth according to Darwinian natural selection.

unitary-self approach concerning the integrity-as-taste view. Nonetheless, Dowell et al., make one single advance. Namely, they model honesty as a discrete choice: either one chooses to take a bribe or not. This modeling resembles the state-dependent utility function [Hirshleifer & Riley, 1992, pp. 60-66] and Robert Moffit’s [1983] model of welfare stigma. However, Dowell et al., place the moral choice in the standard, substantive utility function, and focus on accounting for integrity as a lumpy, discontinuous choice. The point, however, about integrity is not mainly that moral choices are discontinuous and, consequently, that the traditional unitary-self views are faulted for missing this fact. There are a host of other phenomena that are discontinuous, but unrelated to honesty, such as choices taken at the extensive margin—as whether to dress warm, take a vacation, or file for divorce. Although Dowell et al., see their model as intermediate between the unitary- and multiple-self views, they miss the point of contention that divides these two approaches. Namely, the contention is not about the continuity versus discontinuity of behavior, but rather is about the continuity versus discontinuity of the self. Dowell et al., chose the unitary-self conception and, hence, their model suffers from the same fate: Why does one usually experience shame when one sells his integrity?
theory.

1.1.3 The integrity-as-trait view

A Darwinian selection argument interestingly can come to the rescue and remedy the viability anomaly facing the integrity-as-taste view [e.g., Witt, 1986; Hirshleifer, 1987; Frank, 1987; Güth & Yaari, 1992; Gintis, 2000]. The argument employs what one may call “social selection” as opposed to familiar natural selection. In social selection, the selectors are conspecifics—i.e., part of the same population as in the case of sexual selection—who favor agents with the trait for honesty over agents with the trait for deception. According to this argument, if honest agents in the population exceed a critical threshold, they would reproduce faster than conniving agents for the simple reason that traders, especially in single-shot games, prefer them over deceivers. One hurdle in the integrity-as-trait view, which is not serious, is that the signal for honesty, such as sweating and voice trembling, should be expensive to mimic.

So, the integrity-as-trait view explains viability. Further, the integrity-as-trait view sheds light on shame: Shame can be seen as the emotional/physiological cost of inhibiting one’s trait to be honest. That is, it can be seen as a “hard-wired” apparatus residing in the organism to back-up the honesty trait.

However, the integrity-as-trait view invokes its own enigma. How can the integrity-as-trait explain self-rationalization? If the agent does not have the integrity trait, why is he pretending to himself that he has it? On the other hand, if the agent does have the integrity trait,
the integrity-as-trait view may explain self-rationalization as the attempt of the agent to avoid the hard-wired pain of shame. This would be an illustration of having the cake and eat it too: Here, the organism is capable of authentically sending the cues of trust, but only to trick others while (thanks to self-rationalization) feeling the pain of shame.

However, why would nature undertake such a circuitous route? Such a route is, in the final analysis, no different from mimicking the physiological cues of having the trait of integrity, when the agent does not. If this is the case, it would be cheaper and, hence, more efficient for nature to produce agents who mimic the trait of integrity than, first, to produce agents with the trait of integrity but only, second, to undermine it with the trait of self-rationalization.

In addition, the idea of shame as a “hard-wired” apparatus is doubtful to start with. It does not explain why agents, in many instances, can suspend the shame apparatus without resorting to self-rationalization. As the model below shows, agents find it suitable to negate integrity, without arousing shame, under particular circumstances. For instance, if agents become convinced that the only way to survive is through cheating, they do not have to worry about commitments to the past self—and they would cheat without arousing any shame. So, shame is not an automatic, “hard-wired” response. Shame, as shown in the model, can be turned on or suspended—depending on the current circumstances.

Another anomaly faces the integrity-as-trait view: The diffusion of the honesty trait can take place only if the population has already reached a critical ratio of honest agents. But what is the origin of such initial ration? So, the integrity-as-trait cannot, in the final analysis explain the origin of honesty.
So, the puzzle persists. The three flavors of the unitary-self approach cannot easily explain integrity, shame, or self-rationalization.

1.2 Multiple-Self Approach

The multiple-self approach seems to account for integrity, shame and, correspondingly, for self-rationalization. The multiple-self approach comes in different varieties [e.g., passim Elster, 1986]. What concern us here are the multiple-self theories related to moral behavior. For these theories, Immanuel Kant’s [1969] notion of the “categorical imperative” serves as the basis of the idea that a moral self stands independently from substantive utility. For Kant, the categorical imperative amounts to acting according to duty as dictated by the maxim that one acts according to rules that he wishes them to be universal laws. So, the maxim does not hinge on one’s preferences or particular goals. For Kant, moral principle can only stem from the categorical imperative that universalizes action. As such, there must be a moral self that is independent of the substantive utility that animates self-interest.

Following Kant, Amitai Etzioni [1986] argues that one should distinguish interests, which generate "pleasure utility," from ideal ends which engender "moral `utility'." The two kinds of utility are supposedly incommensurable—a position also advocated by a number of economists critical of the unitary-self approach [e.g., Harsanyi, 1955; Hirschman, 1985].

Amartya Sen [1977, 1995] takes the argument a step further. He regards duty, or what he calls commitment, to be “counter preferential” in the sense that it has a higher moral source that is bound to reduce welfare or what is called here substantive utility. Sen welcomes John
Harsanyi's [1955] distinction between "ethical" and "subjective" preferences. While the former expresses what the person prefers "on the basis of impersonal social considerations alone," the latter denotes what the person actually prefers "on the basis of his personal interests or on any other basis" [Ibid., p. 315]. This Kantian dichotomy draws a wide and unbridgeable distinction, as if the social "good" is metaphysically separate from individual welfare. Sen [1977] enriches this dichotomy by proposing the notion of meta-preferences [see also Frankfurt, 1971]. For Sen, the structure of the self is more complex, where the social good can be stratified into multiple levels, ranked by a higher principle of morality.⁶

One can state in general that the multiple-self approaches are successful in explaining shame. However, the explanation gives rise to another set of anomalies. First of all, how does the moral self decide what is the moral thing to do? If such self is totally independent of interest, can it be trusted to determine the moral action—given that action is ultimately about interest? Second, what is the relation of the moral self with substantive utility? If agents truly have multiple selves, are they suffering from the multiple personality disorder? Or more seriously, how can one bridge among the multiple selves if the agent is not at the extreme state of disorder? This bridging is almost impossible given that the spring of the moral self is thought to be independent

⁶ Commenting on Sen, Jane Mansbridge [1998] regards such Kantian approach too restrictive because is treats altruism (sympathy), or what she and Elster [1990] call “love,” as part of substantive utility (welfare or interest). For Mansbridge, Kantian duty and love should be seen as independent springs, and in turn neither can be reduced to substantive utility.
from substantive utility. Third, the phenomenon of “identity switch,” as in the case of spiritual
rebirth, is anomalous for the multiple-self approach. Fourth, and of most importance, if moral duty
is unrelated to interest, it would give rise to sentimental fools, i.e., people who act according to duty
when they are excused not to. For instance, it would be stupid to be a whistle blower, and lose one’s
job, if there is no one in the community who cares. However, if we take into consideration heroism,
as done at the end, it might not be a foolish act.

The proposed quantum view promises to provide an account of integrity that supersedes the
shortcomings facing both the unitary-self view as well as the multiple-self view. But first we need to
define the scope of the investigation.

2. Conceptual Issues

2.1 Utility: Symbolic vs. Substantive

It is proposed here that there are two kind of utility, “symbolic utility” and “substantive utility.”
They are defined in the following way:

Definition: Substantive utility is the ordinary satisfaction (welfare) arising
from the consumption of shelter, food, objects of art, and so on.

Definition: Symbolic utility is the non-ordinary satisfaction arising from
the context of the consumption of substantive goods that make up the
substantive utility set.
The proposed distinction is not a dichotomy as in the case of the multiple-self approach. The connection between the two kinds of utility can be expressed in three ways:

1. **Text/Context:** Substantive utility relates to symbolic utility in the same way that a word relates to the meaning or context afforded by the whole sentence. For example: Examine the word “heirloom” in the following two sentences:

   1) “This is an heirloom from her grandmother”
   2) “This is an heirloom tomato from her grandmother”

2. **Subset/Set:** Substantive utility relates to symbolic utility in the same way that a subset relates to the set. The set cannot be a subset of itself. Otherwise, one commits The Cretan Liar Paradox

3. **Ordinary/Quantum:** Substantive utility is ordinary while symbolic utility is quantum. The term “quantum” is used in two related senses:

   1) Symbolic utility expresses a unity of satisfaction that cannot be divided;
   2) Symbolic utility expresses a quality of satisfaction that is neither totally separable from substantive utility, nor is an ordinary element in the substantive utility set.
2.2 Symbolic Utility: Integrity vs. Respect

Shame differs from embarrassment. Shame is the emotion that may arise from violating the symbolic utility of integrity (as when one acts with dishonesty). Embarrassment is the emotion that may arise from failing to be forward-looking sufficiently, more tenacious, or more alert in pursuing one’s goals. Such failure negates another kind of symbolic utility called here “respect utility” [see Khalil, 2003c]. While shame is close to “foolishness” in the sense that one did not \textit{ex ante} choose the optimum choice, embarrassment differs from “foolishness.” As Arnold Buss notes:

Unlike embarrassment, in shame there is usually a stricken look, often one of sadness. People verbalize feelings of regret, chagrin, and humiliation, for they are mortified at their own behavior and have a strong desire to escape from the eyes of onlookers or a desperate wish to sink into the earth. There is none of the feeling of \textit{foolishness} present in embarrassment. Rather, shame involves sorrow, even self-disgust. Consider these quoted reactions to a shameful event reported by college students: Wanted to crawl under a rock and die; felt dumb and humiliated because everyone knew; felt hurt, confused, regretting; wanted to cry and try to hide it from others; felt less of a person than others; depressed, mad at myself; felt failure, humiliation, regret; and felt ashamed, stupid, and mad at myself [Buss, 2001, pp. 177-178] (emphasis added).

To note, social psychologists have yet to offer a theoretical account of the difference between shame and embarrassment. At best, Buss [2001, p. 178] asserts, without elaborating, that “[u]nlike embarrassment, shame is often the consequence of immorality.” But studies after studies merely focus on the minute and behavioral differences between the two [Miller & Tangney, 1994; Babcock & Samini, 1990]. They focus on scenarios of shame and embarrassment to record which is more serious, involve intentionality, reactions such as blushing, duration of feelings, reaction of others, and so on.
But if we take the observation of Buss seriously, we should ask why shame involves morality, while embarrassment arises from failing to act with attentiveness or apply one’s best effort. As Buss [2001, p. 184] states, in “embarrassment, one feel foolish; in shame, humiliated.”

Embarrassment arises from either social gaffes, when observers are necessary, or from failure to apply one’s best effort, when observers are not necessary [Khalil, 2003c]. In either case, the act is not censured as immoral. In contrast, shame arises from violation of trust or what one takes dear to self-identity. Shame involves immorality. In this regard, it is helpful to distinguish two kinds of commitment, “non-binding commitment” and “binding commitment” [Khalil, 1999].

The term “binding commitment” denotes the promise to pay one’s debt or to honor contract—a promise whose violation justifies punishment or ostracism. Such a promise is non-negotiable, unless the agent is experiencing an emergency situation. Such a promise differs from “non-binding commitment” such as the aspiration to become a singer or the ideal to eliminate poverty [see Khalil, 1999]. Aspirations and ideals are negotiable commitments if their pursuit turned out to be more difficult than expected. The abandonment of the aspiration or ideal does not call for punishment or ostracism. The negation of non-binding commitments at the first sight of difficulty, though, undermines self-respect and invites embarrassment because it indicates that the agent lacks tenacity [see Khalil, 2003c].

The object of study here is binding commitment.\(^7\) Binding commitment invokes a

\(^7\) In either kind of commitment, the phenomenon under study, viz., integrity, has nothing to do with altruism [Khalil, 2003b]. Many authors [e.g., Posner, 2004] offer altruism as an explanatory of honesty. Altruism is primarily a substantive taste towards others, while
particular context, viz., the commitment to the past self. Such a context is called “category” in order to distinguish it from another kind of context, called “frame,” invoked by the non-binding commitment of aspiration, which is forward-looking.

2.3 Object of Evaluation (Integrity Utility): Action vs. Self

While the focus here is on shame, we need to distinguish between two kinds of evaluations that give rise to shame. Shame can arise from the evaluation of the self in toto, but also it can arise from the evaluation of an action. The evaluation of the total self involves, if negative, humiliation or ignominy [Miller, 1993]. The focus of this essay is on the evaluation of a specific action. This should not mean that evaluation of total self is insignificant.

As Figure 1 shows along the vertical axis, these two kinds of self-disapproval appear if we specify the object of evaluation. If the object is the total self, i.e., there is no specific action which is under evaluation, the disapproval is either humiliation (when public exposure is necessary) or ignominy (when public exposure is unnecessary). Given the focus on binding action (not embarrassment), this paper ignores humiliation and ignominy.

2.4 Shame (Integrity Utility): Opportunism vs. Ridicule

While the focus here is on shame (not embarrassment) in relation to the evaluation of a specific action (not total self), we still have two kinds of shame that are occasioned by the issue of whether public exposure plays the central role. As Figure 1 shows along the horizontal axis, if commitment is primarily motivated by the satisfaction of one’s self-image.
public exposure is necessary, it gives rise to “shame of ridicule.” But if public exposure is not necessary, but definitely would make self-disapproval worse, it gives origin to what is coined here “shame of opportunism.”

The choice of terms may seem idiosyncratic, but hopefully they correspond to some extent to everyday usage. An alternative terminology [Tangney et al., 1994], not followed here, is “private shame” vs. “public shame.” Private shame arises from an opportunistic act. It does not depend on whether others are aware of the act—although such awareness intensifies the emotion. In contrast, public shame, when it would be absent when no one is watching, arises from that which necessarily entails the presence of onlookers, as explained shortly.

Shame of opportunism, the main focus of this paper, arises from weakness of will that hinders one from completing (binding) commitments—commitments that would detract from what is named here “substantive utility.” Shame of opportunism emerges when one behaves as a free-rider such as when one acts opportunistically and reneges on a promise made to others or to the self. The term “substantive” is used to denote material interest or welfare-enhancing tastes such as shelter, food, and aesthetic pleasure. To make sense of substantive utility, it stands juxtaposed to self-regarding emotions, what was called above “symbolic utility,” that encompasses self-integrity and other quantum effects ranging from self-respect to the expression of admiration [Khalil, 2000]. Given that substantive utility is about direct, non-mediated welfare, shame of opportunism, or self-rationalization, does not need the presence of others—although such presence heightens the shame.
On the other hand, shame of ridicule stems from failing to adopt another kind of (binding) actions, viz., what enhances what is explained below as “reflexive utility.” Shame of ridicule consists of committing indecent exposure or acts that violate identity roles that can be deliberate or out of sloth. They consist of obvious violation of rules that respect the “property rights” of others—where such rights are not in the substantive sense but rather in terms of sensitivities. For instance, shame of ridicule arises when the agent exposes in public, oblivious to the feelings of others, certain body parts, when he violates dinner table manners, and when he gives a gift that would ridicule the recipient in public. Of course, the particularities of indecent exposures vary culturally; they could be belly buttons or sexual organs.

The term “reflexive” is chosen because the indecency concerns utility (either of the substantive or symbolic kind) that is invoked by association or in an indirect manner. It tarnishes the original utility indirectly. The original utility under focus can be substantive, as in the case of indecent exposure, or symbolic, as in the case of the gift. While original utility can be substantive or symbolic, reflexive utility is whatever extra pain gained as a result of the pain of others. For instance, the exposure of body part may make others repulsed, which affects the disposition of everyone. Or, in the case of the gift, the mortification of the recipient of ridiculous gift increases the giver’s taste for revenge. As these examples indicate, the presence of others is, unlike shame of opportunism, necessary.

As shown in Figure 1, what is common to shame of ridicule and humiliation is the presence of others—a common characteristic that may explain why some authors confuse the two. We should not lose sight that what distinguishes them is the object of evaluation. In ridicule, a
particular act is ridiculed. But in humiliation, the core identity of the person is judged, such as when one makes fun of someone’s ethnicity, religion, accent, parents, posture, weight, nationality, physical appearance, and physical or mental disability. Given the focus here on shame, the object of evaluation is (binding) action rather than the total self.

3. A Simple Model of Integrity

3.1 Three Scenarios of Trespassing

Violation of the law by itself is not a cause of shame. Cheating on income tax or ignoring minimum wage laws usually does not cause one to feel self-disgust. On the other hand, a diverse number of moral codes which are at the origin of stigma, such as the ones concerning virginity in many cultures, have no correspondence in law. But in other cases, law and morality may overlap, as in the violation of property rights as in the cases embezzlement, espionage, and theft.

What concerns us here is the moral code rather than the law per se. I want to use a simple example of the violation of the moral code with regard to trespassing inspired by Ronald Coase’s [1960] efficiency theory of property rights. I want to use three scenarios of trespassing in order to anchor the proposed modeling of integrity qua quantum utility:

Example: The perpetrator, P, without permission, lets his cattle march through the property of the aggrieved party, A. The property is posted throughout with "no trespassing" signs. While A’s wealth decreases by $\Delta W_a$, P’s wealth increases by $\Delta W_p$.

There are three possible scenarios:

Scenario 1: $\Delta W_a > \Delta W_p$ So, it is inefficient according to Coase Theorem.

Scenario 2: $\Delta W_a < \Delta W_p$ – when there is no “emergency” involved. So,
while it is efficient according to Coase Theorem, it does not qualify under the principle of imminent domain.

**Scenario 3:** $\Delta W_a < \Delta W_p$— whereas the violation of A’s right was undertaken because of emergency. So, while it is efficient according to Coase Theorem, it falls within the principle of imminent domain.

### 3.2 The Quantum Model

The proposed model promises to sharpen the issues involved in the three scenarios. Let us assume that the perpetrator (P) maximizes his substantive utility function ($U^s$),

$$\text{Max } U^s = U^s(X_1, X_2, \ldots, X_n) \ldots (1)$$

where $X_i$ ($i = 1, \ldots, n$) denote substantive goods such as food, shelter, museum visit, and so on. The usual assumptions concerning the $U$ function ($U' \geq 0$ and $U'' < 0$) apply. The agent $P$ maximizes his consumption subject to two constraints: “the efficiency constraint” and the “ethical constraint.” Concerning the efficiency constraint,

$$W_p(I) + \Delta W_a = P_1X_1 + P_2X_2 + \ldots + P_nX_n \ldots (2)$$

where $W_p(I)$ expresses perpetrator’s wealth (contingent on integrity (I)), $\Delta W_a$ the change of wealth of the aggrieved party (in case the perpetrator cheats), and $P_i$ the prices of the $X_i$ goods. The usual assumptions ($W_p(I) \geq 0$ and $X_i \geq 0$) apply. Also, it is assumed that the discrete value of I is common knowledge with zero transaction cost. While $\Delta W_a \leq 0$ is the usual case, the conclusions of this model do not change in the unusual case of $\Delta W_a > 0$.

This construction of the efficiency constraint involves the internalization of all externalities. The reason for the internalization has nothing to do with altruism. Otherwise, the utility function should reflect the taste for altruism. Rather, in this model, information is fully
transparent and perpetrators are held accountable at zero transaction cost. That is, whatever the agent imposes on others because of cheating, he would have to compensate the aggrieved party, following Coase Theorem, while the cost of enforcing the law is zero.

The efficiency constraint (2) involves two further assumptions. First, integrity (I) is a category, i.e., a discrete variable that shifts the constraint function. While the quantity of a bribe is continuous, the act of taking or rejecting the bribe is categorical, i.e., discontinuous. Integrity can take only two values: \( I = 0 \) if the agent acts with decency, and \( I = 1 \) if the agent violates his own standard of decency. Second, this model assumes that the perpetrator does not resort to dishonesty for purposes such as retaliation, revenge, or malice. The only purpose is to increase wealth, in the sense of substantive utility, of the perpetrator. So, by definition, the change of wealth is, at least \textit{ex ante}, positive,

\[
\Delta W_p(1) = W_p(1) - W_p(0) \geq 0 \quad (3)
\]

The change of wealth of the aggrieved party \( (\Delta W_a) \) may not equal the change of wealth of the perpetrator \( (\Delta W_p(1)) \):

\[
\Delta W_a = r \Delta W_p(1) \quad (4)
\]

\footnote{It is true that one may try be ambiguous when he takes a bribe, such as camouflage it as a series of small and inconsequential gifts, or when he commits a series of small lies. However, such acts of camouflage amount to self-rationalization and, hence, begs the question: Why does not the agent commit the dishonest act in one strike—which should be cheaper from the standpoint of transaction cost. The attempt to be ambiguous (i.e., self-rationalization) when one acts dishonestly only confirms that integrity is a quantum act. The agent who takes small “gifts,” rather than a one-time bribe, is not consuming dishonesty at the margin—but rather is trying to hide (from himself) the dishonesty altogether.}
where \( r \) is the rate of transformation of the perpetrator’s benefit, arising from dishonesty or trespassing. It is usually the case that \( r < 0 \). (But even if \( r \geq 0 \), it should not change the results of this model.) In the trespasser example, in Scenario 1, \( r < -1 \); i.e., the trespassing is inefficient from the standpoint of Coase Theorem. In Scenarios 2 & 3, \( 0 > r > -1 \); the trespassing is efficient from the standpoint of Coase Theorem.

The rate of transformation \( r \) is not a given,

\[
r = r \left( TC_a \left[ w_a, w_p, R_a \right], W_p/W_a \right) \ldots (5)
\]

where \( TC_a \) represents total cost of the aggrieved party and \( W_p/W_a \) the respective wealth levels. \( TC_a \) is a function of \( w_a \) and \( w_p \), the respective wage rates, and \( R_a \), the lost resources of the aggrieved party. The rate of transformation function (5) captures Pareto optimality and, what is called here, Bentham optimality. For Pareto optimality, \( r(.) \) is a function of the total cost \( (TC_a) \), which, in turn, is a function of opportunity cost measured by the wage rates \( (w_a, w_p) \) and actual lost resources of the aggrieved party \( (R_a) \). In addition, for Bentham optimality, \( r(.) \) is a function of relative wealth \( (W_p/W_a) \), which assumes that interpersonal utility comparison is non-problematic.\(^9\) If so, it would be efficient for the poor perpetrator, as Robin Hood had reasoned, to steal from the rich aggrieved party if the marginal dollar generates greater pleasure for the poor than it does for the rich.

Concerning the other constraint, the ethical constraint, it arises when the binding

\(^9\) Actually, Adam Smith’s [1976, Khalil 1990] notion of “impartial spectator” allows for an agent, who occupies a station equally distanced from altruist and beneficiary, to compare their welfare. Even Gary Becker’s [1981] theory of altruism, based on including the welfare of beneficiary to be part of the utility of the altruist, assumes that interpersonal utility comparison is non-problematic.
commitment to be honest is part of the perpetrator’s integrity utility ($U^i$):

$$\Delta U^i = \Delta U^i (d_j I(S) | W(0)_p > W^c_p) \ldots (6)$$

where $d_j$ ($j = 1, 2, \ldots, m$) is the discrete degree of the offense as assessed by the perpetrator, $S$ the number of spectators including the perpetrator, and $W^c_p$ represents the critical wealth of the perpetrator below which places the perpetrator in a state of emergency. The $\Delta U^i$ function is decreasing, in a quantum manner, in $d_j$ when $I=1$ (dishonesty), i.e., shame arises, given that the wealth of the perpetrator acting honestly ($W(0)_p$) exceeds his critical wealth ($W^c_p$). That is, in emergency situations, if the agent was to act honestly and, consequently, would place his wealth below the critical level, integrity utility would not be affected when $I=1$:

$$\Delta U^i = \Delta U^i (d_j I(1)(S) | W(0)_p < W^c_p) = 0 \ldots (7)$$

The exception to the rule of justice in the emergency zone, where there is imminent danger to livelihood, is long recognized in Roman law as well as in the common law of England and the United States. According to this legal tradition, a contract made under necessity, such as imminent danger to life, is not upheld by the legal authorities [Epstein, 1995, ch. 6]. Timur Kuran [1990, 1995] discussed at length what he calls “preference falsification.” Agents had to resort to dishonest behavior in order to survive in oppressive environments such as the “Grand Inquisitions” in Spain and modern dictatorships. Agents had to lie about their true identity and beliefs. Such lies, when honest wealth would be under the critical level, usually did not cause the agent to incur a sense of shame.

The conditional ethical constraint (7) highlights the difference between Scenario 2 and Scenario 3 in the trespassing example. Although both scenarios enhance efficiency from the
standpoint of Coase Theorem, only Scenario 3 meets the ethical criterion because the perpetrator’s wealth would be below critical wealth if he acts with integrity. Of course, this requires a greater specification of critical wealth.\(^{10}\)

\[
W^c_p = W^c_p (d_j(W_p/W_a)) \quad . . . (8)
\]

That is, \(W^c_p\) is defined in relation to the contemplated degree of the offense. Murder can be justified if enacted in self-defense, but not if enacted to stave off hunger. While shoplifting of food can be justified if enacted to stave off severe hunger, but not if enacted to save money for a theatre ticket.\(^{11}\) So, \(W^c_p\) decreases with \(d_j\). But \(d_j\) is not a given. It is influenced by the wealth of the potential perpetrator relative to the wealth of the aggrieved party. It makes no sense in terms of criticality of wealth for a starving person to steal from another, equally starving person.

To elaborate, the higher \(j\) means greater degree of offence, which is relevant only if \(I=1\) (dishonesty). The \(j\) degree is subjective in the sense that different agents might assign different degrees of offense to the same act of opportunism. However, usually and across the board, murder is a higher degree of offence than taking a bribe, and taking a bribe is a higher degree of offense than shoplifting or failing to declare a defect in one’s used car [Akerlof, 1970].

Furthermore, for \(\Delta U^j\) to be non-zero, \(S \geq 1\), i.e., the self has to be aware of its own action for integrity utility to change. Obviously, \(d_j I\) increases in \(S\), where the pain of shame is amplified by the number of people who are aware of the dishonest act of the perpetrator.

\(^{10}\) Criticality of wealth is not an absolute level of poverty or an ironclad state of emergency. It rather includes what the agent regards, with the support of conventions, what is needed to sustain "capability" as defined by Amartya Sen [1985].

\(^{11}\) Likewise, the eating of the flesh of already dead people, as abhorrent as it could be, was sanctioned by the Catholic church, and was not considered a sin, in the case of the survivors of
To maximize substantive utility ($U^s$), the necessary condition for cheating, but not the sufficient condition, is if net social wealth increases:

$$\Delta W_p(1) + \Delta W_a \geq 0 \ldots (9)$$

If we use equation (4), the necessary condition (8) becomes,

$$\Delta W_p(1) + r\Delta W_p(1) \geq 0 \ldots (10)$$

or what one may name the “optimum efficiency condition”:

**optimum efficiency condition**: $r \geq -1 \ldots (11)$

If one cheats when $r < -1$, the act would be foolish, i.e., non-optimum in the efficiency sense.$^{13}$

However, there is a sufficient condition, what one may name “optimum ethical condition”:

**optimum ethical condition**: $\Delta U^i (I=1) \geq 0 \ldots (12)$

If one cheats ($I=1$), the ethical condition can be satisfied only in cases of emergency. This means that integrity utility is not the same as the act of integrity. This optimum ethical condition amounts to a corner solution similar to the optimum efficiency condition. As much as agents in many occasions act foolishly and violate the optimum efficiency condition, they can also act unethically and violate the optimum ethical condition.$^{14}$

---

$^{12}$ Libertarians criticize Coase Theorem for focusing exclusively on this condition.

$^{13}$ We are assuming full transparency and accountability—so the person would have to compensate fully the aggrieved party’s substantive damages.

$^{14}$ The optimum ethical condition stipulates “no-tradeoff” zone between integrity and interest beyond what is deemed as emergency situation. This differs from lexicographic ordering. While the discontinuous zone or no-cheating rule is absolutely non-tradable beyond critical wealth,
3.3 Discussion

The proposed model of integrity embodies the quantum approach which distinguishes it from the unitary- and multiple-self approaches. Contrary to the unitary-self approach, it upholds the idea of integrity as a symbolic utility that is separate from substantive utility by setting up integrity utility as a distinct function. On the other hand, contrary to the multiple-self approach, it upholds the idea of integrity as a symbolic utility that is somewhat dependent on substantive utility by making the function related to substantive interest. In fact, the other approaches are simply special and limited cases of the quantum view.

3.3.1 The Unitary-Self Approach Revisited

In the view of integrity-as-strategy, there is no ethical constraint (6). There would be only the efficiency constraint (2a),

\[ W_p(I) + \Delta W_a + R = P_1X_1 + P_2X_2 + \ldots + P_nX_n \ldots (2a) \]

Whereas R (R<0) represents rent paid by the “perpetrator” to the “aggrieved” party. Given that the “perpetrator” will not violate the rights of others because of concern over reputation, the “perpetrator” must obtain a permission to use the resource of the “aggrieved” party, besides paying the aggrieved party the full cost of such use (\(\Delta W_a\)). So, when I=1, it means that the “perpetrator” decided to ask for a permission to use the resource of the aggrieved party. In this lexicographic ordering allows for a continuous trading of bundles after a minimal consumption of one commodity (such as basic food) is satisfied.
case, the owner of the resource may charge rent (R) determined by the market or, in the case of bilateral monopoly, by negotiation. The only condition that holds is,

$$\Delta W_p(1) + \Delta W_a + R \geq 0 \ldots (9a)$$

After substituting, using equation (4),

$$\Delta W_p(1) + r\Delta W_p(1) + R \geq 0 \ldots (10a)$$

**optimum efficiency condition with rent**: $r \geq - (1 + R/\Delta W_p(1)) \ldots (11a)$

However, as stated above, while this case can account for integrity in iterative game under particular conditions, it cannot account for integrity in single-shot game with zero transaction costs.

The integrity-as-taste view can account for integrity in single-shot game with zero transaction costs. In the integrity-as-taste view, there is no ethical constraint (6) as well. The efficiency constraint (2), and also the optimum efficiency condition (11), would hold as in the quantum model. The only difference is the utility function of the perpetrator,

$$\text{Max } U^s = U^s (I, X_1, X_2, \ldots, X_n) \ldots (1)$$

where $I=0$ for decency and $I=1$ for dishonesty. Since integrity is a categorical choice, the additional constraint for committing dishonesty, beside the optimum efficiency condition (11) is the following egocentric ethical condition:

**optimum egocentric ethical condition**: $U (\Delta W_p(1) - \Delta W_a) \geq U(I = 0) \ldots (12a)$

It is called “egocentric” because the perpetrator is placing a price tag on his integrity [Khalil, 2001]. For such an egocentric, he is justified in betraying his partner if the utility of the net
income arising from cheating exceeds, or at least equals, the utility arising from acting with integrity.

However, as stated above, this view cannot account for the persistence of shame, where at least the non-egocentric agent does not show pride in abiding by the above constraint (12a). Also, such a taste for integrity makes the organism non-viable in competitive markets.

The integrity-as-trait view attempts to take care of the issue of viability and persistence of shame. In the integrity-as-taste view, there is no ethical constraint (6) as well. The agent here acts with integrity because he is built that way—and hence shame persists because it is hard-wired. The agent is built that way because, according to the Darwinian social selection story, other agents prefer to deal with a person who has the integrity trait. Agents with the integrity trait, under particular conditions, would be richer on average than deceitful agents:

$$\sum_{t} \delta^{t} W^i_p(0) \geq \sum_{t} \delta^{t} W^i_p(1) \quad (2b)$$

where $\delta < 1$ is the discount factor, $t = 0, \ldots, \infty$, assuming that the agent also counts the benefits of the integrity trait for potential progenitors.

However, for the integrity-as-trait story to work, there has to be a critical mass of agents with the integrity trait. This defeats the purpose of explaining the existence of the trait. Further, as mentioned earlier, if integrity is a trait, why do people violate it and then try to rationalize that they did not? The phenomenon of self-rationalization is enigmatic in the house of the integrity-as-trait view. The agent who does not have the trait need not pretend to himself that he has it. The agent who has the trait, but undertake self-rationalization when he violates his trait, would be a strange, inefficient organism. First, the organism is loaded with the integrity trait, then he is
loaded with self-rationalization in order to avoid the sense of shame. It would be more efficient for nature to build, instead, an organism with the dishonest trait accompanied with the ability to mimic the honesty trait in order to free-ride.

3.3.2 The Multiple-Self Approach Revisited

To avoid all the anomalies facing the unitary-self views, one is prompted to start afresh, as many of the advocates of the multiple-self approach do. Unlike the three flavors of the unitary-self view, the multiple-self approach recognizes the ethical constraint,

\[ \Delta U_i^j = \Delta U_i^j (d_j I(S) \Psi W(0)_p) \ldots (6a) \]

Here, the multiple-self view identifies integrity utility with the act of integrity itself. It does not allow for a zone of emergency where, if honest wealth is below the critical level, one would be allowed to cheat or to abrogate an earlier commitment without incurring shame.

The neglect of this zone of emergency is not accidental. The advocates of the multiple-self approach do not seem to realize the origin of binding commitment, viz., as part of production activity. The binding commitment is originally undertaken as an investment in capital goods. It expresses the sense of bonding that allows agents to connect their present self to their previous self. Such a connection is needed in order to organize one’s past skills, as discussed below.

As mentioned earlier, the multiple-self approach leaves the agent hanging with a taste for principle without connecting it to substantive well-being. It cannot explain why agents switch from one identity to another. That is, it cannot afford an endogenous account of the development of identity, i.e., specifying the set of incentives that propel one to switch his core identity. By
radically separating integrity utility from substantive utility, the advocates of the multiple-self view cannot easily connect them back.

4. Ramifications

4.1 Identity Switch

Short of emergency, it is hard to satisfy the optimal ethical condition. Short of emergency, as examined above, agents usually resort to self-rationalization, i.e., the over-stretching of a narrative in order to justify a dishonest action. The mechanism of self-rationalization attempts to capture integrity while the agent is acting dishonestly.

However, one may still revoke binding commitments, even when one is outside the emergency zone, and still avoid the sense of shame without resorting to self-rationalization. This is possible if one undergoes “identity switch.” Unlike the case of justified dishonesty in the case of emergency, in identity switch the agent no longer recognizes the old rules of integrity, morality, or identity. In identity switch the agent preserves integrity while disowning one’s binding commitments.

Identity switch is problematic for the multiple-self view. Why should one undergo identity switch if he has already co-existing multiple selves? The switch from one core identity to another, from a conservative to a liberal or from a Catholic to a Jew, is usually undertaken as a result of an identity crisis. The agent might be seeking a greater coherence or comprehensibility—as probably the case with St. Paul’s spiritual rebirth in the New Testament. Or the identity switch can be undertaken because of expected wealth. It is often the case that
agents switch their religions in light of the religion of the ruling class or the religion of the
hegemonic state. In addition, when states change directions in revolutionary times, such as
attacking old friends while embracing old enemies, they undergo identity switch. Further, agents
may not openly undergo identity switch but rather slowly slip into self-hate and even group self-
hate [Gilman, 1986].

4.2 Imprinting and the Transactional View

The idea of imprinting originated from the work of Konrad Lorenz [1965] and other ethologists.
They observed that organisms have a great partiality towards whoever is their first care provider
or their first physical surrounding. They exhibit what behavioral economists call the
“endowment effect” [Thaler, 1980], which is marshaled to explain an observed difference
between the "willingness to pay" and "willingness to accept" [Camerer, 1995, pp. 665-670;
Eisenberger & Weber, 1995]. The agent has a bias towards the inherited endowment or default
state because the first provider or first physical surrounding acts as the solid rock that the agent
needs in order to produce,

\[ Q = Q(K, L, T, U^i) \ldots (13) \]

where Q is output, K capital, L labor, and T land. In this production function, integrity utility
acts as “identity capital” that affords the agent “attention.” Attention includes memory and focus
that are needed to coordinate human capital with other physical resources over time. The fund
\[ U^i \] is simply the aggregation of all previous additions of integrity arising from previous, discrete
choices,
\[ U^i = \sum_t \Delta U^i_t \ldots (14) \]

where \( t = 0, \ldots, T \), \( T \) the present, and \( \Delta U^i_t \) is the ethical constraint (6). In this manner, the agent’s productive ability, the ability to undertake forward-looking action, depends not only on the traditional inputs but also on amassed fund of integrity utility.

Imprinting in human affairs amounts to attachments to the binding commitments made by past selves. The agent invests in honesty or authenticity to augment identity capital as he invests in other resources such as human capital. In the case of integrity, the agent is relating to the first self as the solid rock, the first care provider, upon whom he undertakes further explorations and production activity. So, past commitments act as the default states which are not initially questioned. Norms of justice and binding commitments play analytical roles similar to the role of the first care provider in imprinting. One should not, and usually does not, start from scratch or from some an idealized world [Levi, 2004]. Rather, and this is the core of the idea of action according to the transactional view of John Dewey and Arthur Bentley [1999], the agent takes what is given without reflection. The basic idea of the transactional view is that the agent questions current commitments, understanding, identity, and so on, only in light of a crisis or disturbance that needs to be settled [Ryan, 2002].

4.3 Heroism
Heroism is usually confused with integrity [e.g., Urmson, 1958; Heyd, 1982; Montague, 1992]. While they overlap, they are not identical. Heroism can be defined as acting honestly or according to identity when the consequent wealth is below critical wealth:

\[ W(0)_p < W^c_p \]

That is, heroism amounts to the preservation of identity or integrity when the agent, given that he is in a state of emergency, should abandon his integrity. So, in the act of heroism or supererogation, the agent is acting foolishly, not observing the optimum ethical constraint.

Nonetheless, we still admire the thousands of members of Jehovah's Witnesses who chose to perish in Nazi death camps rather than gain freedom by signing the declaration card which renounces their faith and pledges allegiance to the Third Reich [Engardio, 1996]. It is curious how the choice of death or loss of career over the renouncement of one's identity or belief—as also evident by the sacrifices of early Christians, liberation rebels, revolutionaries under repressive regimes, or Samurai warriors\(^{15}\)—is usually admired.\(^{16}\) In fact, every state,

\(^{15}\) Saikaku Ihara's [1981] account of giri (the code of honor among the Samurai) makes an interesting and pertinent reading. Writing in 1688, Ihara's tales are permeated with social satire of the absurdities generated by extreme and unintelligent pursuit of honor. While some tales portray honor as optimum and hence admirable, the majority of tales show how giri engenders suboptimum ridiculous results when the actor pursues giri while he is in the critical zone.

\(^{16}\) One cannot but admire Vaclav Havel when he explained later why he risked so much by sending an open letter critical of the Communist government of Czechoslovakia in 1975:

> The letter, on the primary level, was a kind of autotherapy: I had no idea what would happen next, but it was worth the risk. I regained my balance and my self-confidence. I felt I could stand up straight again, and that no one could accuse me any longer of not doing anything, of just looking on in silence at the miserable state of affairs. I could breathe more easily because I had not tried to stifle the truth inside me. I had stopped waiting for the world to improve and exercised my right to intervene in that world, or at least to
ideology, and religion has its legends of heroes and saints.

If heroism is not foolishness, how should we model heroism? Evidently, the hero’s critical wealth is much lower than the normal critical wealth of spectators. In fact, the hero seems to be most interested in the identity of such spectators. Early Christians withstood torture and met death, rather than abandon their faith, in order to influence the faith of spectators. There is no hero without admirers or potential admirers. What would happen to the Catholic Church if all Catholics become celibate priests and nuns? So, heroism is not an optimal or rational strategy for all the members of the group. Rather, the hero does want everyone in the group to follow his steps to martyrdom. He rather wants others to remember him posthumously. The hero wants to influence the identity of the admirers, taking into consideration the expected utility of the adhered faith or identity.

4.4 Procrastination as Self-Cheating

In all acts of cheating, the present self does not honor the commitment made by a past self. While in cheating proper the aggrieved party is another person, in self-cheating the aggrieved party is a future self.

The best example of self-cheating is procrastination. The self at time $t=0$ promises to cut the grass later, let us say $t=1$, but at $t=1$ the self cheats and does not carry the commitment, imposing a cost on the future self, the self at $t=2$. This imposes an extra cost because the self at $t=2$, as understood at $t=1$, is already crowded with other responsibilities. Procrastination is express my opinion about it [quoted in Kuran, 1995, p. 287].
usually not seen as dishonesty because the aggrieved party is still the same person. Another example of self-cheating is anorexia [Bell, 1985; Condit, 1990], where the present self undermines the well-being of the future self.

Self-cheating has given rise to pre-commitment contracts, where the present self limits its choice to preserve the interest of the future self. Similar to the inter-person exchange of hostages to prevent defection, pre-commitment is intra-person exchange of hostages to prevent self-cheating. The best example of pre-commitment is the fable of Ulysses and the Sirens [Elster, 1984]. As Richard Posner [1995] suggests, the role of law in molding personal choices can be seen as pre-commitment. For instance, the social security system can be seen as a freely appointed guardian against undersaving, similar to the helmet law in some states.17

4.5 The Penal Code

While the debate in economics on how to model integrity is recent, the debate in legal and political philosophy extends at least to Socrates. It has taken a new form in modern positions as best exemplified in the Kantian orientation of Ronald Dworkin [1977, 1978] and the

17 To caution, however, not all pre-commitments are about the prevention of self-cheating. In many cases, the pre-commitment is undertaken because one admits that he lacks the resolve or tenacity to execute a task. In other cases, the precommitment is undertaken to galvanize resolve as the case of an invading army burning its own ships. Such kinds of precommitment are about forward-looking action, which is outside the focus here.

Furthermore, the proposed quantum view of integrity suggests that a penal code which exclusively relies on depriving the criminal of substantive utility—a code advocated by Jeremy Bentham on the bases of deterrence and retributivism—is inadequate. Some legal scholars, such as Dan Kahan [1996], want to revive the expressive theory of punishment which prevailed before Bentham. The expressive theory entails the appropriateness of the use of shame—inflicted on convicted criminals through corporal punishment ranging from public display of the offense to chain gangs—as a method of punishment.

To wit, as Kahan argues, the penal code of the U.S. is totally detached from the use of shame—as evident by the public's endorsement of imprisonment over cheaper and probably more effective methods of sanctions such as fines and community service. The overwhelming public endorsement of imprisonment as the prime method of punishment does not arise, as Kahan maintains, from misinformation, lack of knowledge, or belief in the greater effectiveness of imprisonment. The public rather regards these alternatives as falling short of expressing the condemnation of acts which violate public trust—a condemnation which goes beyond the demand of substantive retribution.

4.6 Etiquettes

People who are committed to the rights of others usually follow, even in single-shot games, the
rules of fairness in order to reach equitable distribution of public goods, such as access to parks or roads. When an equitable access is not clear, people act with courtesy or what is called generally “etiquettes.” Etiquettes include the rule of “first-come-first-served” when people compete for the services of a clerk at a post office. They also include courtesy and yielding when a person asks if others are interested in the last serving of a dish at dinner table before the person helps himself to it.

So, etiquettes are methods to distribute nonexclusive, public goods which may experience congestion. Etiquettes express binding commitments to fairness; they generally do not express sympathy with the substantive interest of the other. Amartya Sen expresses the difference between commitment and sympathy in the following famous parable:

The contrast between sympathy and commitment may be illustrated with the story of two boys who find two apples, one large, one small. Boy A tells boy B, ‘You choose’. B immediately picks the larger apple. A is upset and permits himself the remark that this was grossly unfair. ‘Why?’ asks B. ‘Which one would you have chosen, if you were to choose rather than me?’ ‘The smaller one, of course’, A replies. B is now triumphant: ‘Then what are you complaining about? That's the one you've got!’ B certainly wins this round of argument, but in fact A would have lost nothing from B's choice had his own hypothetical choice of the smaller apple been based on sympathy as opposed to commitment. A's anger indicates that this was probably not the case [Sen, 1982, p. 93; Anand, 1993, p. 67].

To note, etiquettes defined as fairness radically differ from manners such as refraining from telling gross jokes or burping at dinner table. Manners are related to what was called earlier “reflexive utility,” where one’s utility depends exclusively on the utility of others.
Manners in this regard give rise to propriety of behavior or the “virtue of self-command,” according to Adam Smith [1976; Khalil, 1990]. Self-command amounts to restraining the burst of joy or anger beyond what is deemed proportionate to the stimulus. The determination of what is proportionate depends on how other spectators react to the same stimulus.

4.7 Taboos

The proposed theory can open new vistas concerning the modeling of self-identity and taboo markets such as the sale of children, the secrets of one's country, ethnic identity, masculine identity, and so on. The proposed quantum view may answer why adding material incentives does not usually increase, and even may decrease, the supply of taboo items such as cadaveric organs [see Bowden & Hull, 1993]. The shame arising from the trade of taboo items is similar to the shame associated with cheating.

5. Conclusion

The proposed quantum approach neither conceives integrity as an element of substantive utility (unitary-self approach) nor as totally independent of substantive utility (multiple-self approach). The advanced quantum view promises to solve the shame and self-rationalization anomalies that face the unitary-self approach. It also promises to solve the justified dishonesty and identity switch anomalies that face the multiple-self approach.

The paper attempts to solve the puzzles of shame, self-rationalization, justified dishonesty, and identity switch. The key idea is that integrity, which expressed a commitment
to a past self, is a quantum taste. To be a quantum taste, the utility is symbolic rather than substantive. To be a symbolic utility, it expresses the sense of selfhood rather than simply the pleasure of substantive utility. As such, symbolic or quantum utility is neither, under particular conditions, fungible with substantive utility nor it is totally separable from substantive utility.
References


_______. "Rewarding Trust: An Experimental Study." Discussion paper, Europa-Universität Viadrina, Frankfurt (Oder), 1996.


Tokyo: Monumenta Nipponica, Sophia University, (1688) 1981.


