Modeling employees' deviant behavior and employers' reactions: an interdisciplinary approach using principal-agent and prospect theories

Aharon Tziner\(^1\) • Erich C. Fein\(^2\)

\(^1\)Schools of Behavioral Sciences & Business Administration, Netanya University College, Israel
\(^2\)School of Business and Law, Central Queensland University, Australia

Received: 27 July 2012
Revised: 11 November 2012
Accepted: 18 November 2012

Abstract
In this paper we introduce mathematical models to address deviant behavior theory and practice. We examine only employee behaviors and employer reactions, but this framework is innovative and parsimonious, in that it models employer responses to employee’s behavioral deviance based on the dual application of principal-agent theory and prospect theory. After explaining the model, we examine boundary conditions and limitations and propose a series of applications to illustrate the potential usefulness of the model.

Keywords: deviance, deviant behavior, employee economic utility, principal-agent theory, prospect theory
JEL Classification Codes: M52, M21, J33, D20, D03

1. Introduction
At present, workplace deviance has become an important issue and is gaining increasing research attention (Berry, Ones and Sackett, 2007; Cohen-Charagh and Mueller, 2007; and Dilchert, Ones, Davis and Rostow, 2007). The effects of deviant behaviors in the organization have psychological, sociological, economic and managerial implications (Vardi and Weitz, 2004). For example, the financial cost resulting from theft by employees in the United States is estimated at 50 billion dollars per year (Coffin, 2003). Moreover, employees who displayed such deviant behaviors have a greater tendency to resign and develop stress related problems and low morale (O'Leary-Kelly, Griffin and Glew, 1996). They also tend to experience low self esteem, an increase in anxiety and a lack of confidence at work, as well as physical and psychological pain (Griffin, O'Leary and Collins, 1998). Nevertheless, deviant workplace
behavior can also have positive repercussions such as violation of organizational norms that leads to innovation and creativity (Bodankin and Tziner, 2009). However, in this model we shall restrict our investigation to destructive deviance.

2. Modeling deviance calculations in organizational theory research

Robinson and Bennett (1995) define destructive deviance as an intentional behavior that violates significant organizational norms, thus threatening the wellbeing of an organization, its members, or both. Vardi and Weiner (1996) refine the definition by stipulating that the concept refers to behaviors causing harm or having the potential of causing harm that violate societal norms of proper conduct as well. This behavior can be divided into two main categories according to its objective: behaviors that are directed towards other individuals and those that are directed towards the organization. The first category, interpersonal destructive deviance, comprises behaviors such as harassing other employees, stealing from other employees and informing on them. The second, organizational destructive deviance, comprises behaviors such as stealing from the company and sabotaging equipment.

However, at present all the publications on destructive workplace deviance deal only with the conceptual aspects, and the antecedents and outcomes of this phenomenon. To the best of our knowledge, no study has attempted to construct a mathematical model which would enable employers to decide when destructive workplace deviance exceeds the financial value of the deviant worker’s performance. Namely, we investigate situations where the loss due to destructive deviance exceeds the benefits deriving from the deviant’s performance and competence. Such a tool could facilitate the employer in making economically grounded decisions regarding the continuation or discontinuation of a worker’s employment. Furthermore, no attempt has been made to examine the relationship between employees' deviant behavior and the change in the wealth of the employer, whether in a given country or from an international perspective.

Part of the theoretical framework in this proposal is based on Prospect Theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) on one hand, and Contract Theory and Agency Theory (Jensen and Meckling, 1976) on the other. The behavioral approach employed here is based on the fact that all organizations are managed by human beings, and therefore their decisions are likely to involve psychological biases (see, e.g., Devers, Wiseman and Holmes, 2007). Additionally, we focus on deviant behaviors that are based on rational decisions and choices made by an employee (e.g., Vardi and Weitz, 2002). Such misbehaviors may be intended to benefit the self or to inflict some kind of damage (Vardi and Wiener, 1996).

3. Model objectives

The objective of this proposal is threefold: (1) to introduce a mathematical model to fill the lacuna in deviant behavior theory and practice, (2) to provide employers with an accurate framework of judgment and decision making concerning dismissing or keeping deviant employees, and (3) to quantify empirically the relationship between employees’ deviant behavior and the wealth of the employer.

Such a tool would assist the employer in making economically grounded decisions regarding continuation or discontinuation of a worker’s employment. We further explicate how the usefulness and limitations of the model can be realized to cast new light on the dynamics guiding the process of action and reaction of the employee-employer parties engaged in managing deviance at work.
In the next section we present elements of the deviant employee model based on contract theory and agency theory. In the following section we introduce a behavioral model of destructive behavior based on Kahneman and Tversky’s Prospect theory. As with any theory we explain the conditions under which the model is presumed to function. We then discuss limitations, theoretical and practical applications, and next steps in the development of the model.

3.1 Principal-agent components
This research proposal hypothesizes that the decision of an employer to terminate its engagement with an employee, or that employee’s own decision to quit is dependent upon the following variables: the level of an explicit compensation package, the effect of an employee’s deviant behavior, his or her self perceived quality, his or her quality as perceived by the employer, and the replacement costs. Specifically, based on the proposed model, there may be an optimal tolerable level of deviant behavior that any change from this level may result in termination or voluntary quit. In other words, deviant behavior and the employer’s policy are dependent upon a “dead weight” loss generated by deviant behavior. In this model the deviant behavior may very well be a rational decision on the part of the employee.

In this section we will lay out the general principles and assumptions of the model (merging Contract Theory and Agency Theory as in Jensen and Meckling’s (1976) and Fama’s (1980) framework. This leads to the following assumptions.

**Assumption 1.** In a partial equilibrium model, the employer is assumed to maximize the difference between the value of an employee’s contribution \( Q \), (the present value of his or her entire value of contribution over the expected engagement period) and his or her costs, which includes his or her explicit compensation package \( W(Q) \) (the present value of the compensation package over the entire expected engagement period) and the costs generated by potential deviant behavior \( a(W,Q_p)W \) where \( Q_p \) is the self-perceived quality of the employee and \( a() \) is the fraction of costs out of the total value of the compensation package, \( W \).

**Assumption 2.** The employee is assumed to generate a value of \( Q \), but he or she perceives his or contribution as \( Q_p \). While the level of compensation package is \( \bar{W}(Q) \) set at the beginning of the period, his or her desired level is \( W(Q_p) \). Thus, the employee is engaged in deviant behavior attempting to compensate at least partially for the difference between the set level of compensation and his or her desired level.

**Assumption 3.** The key assumption is that an employee’s deviant behavior generates benefits for the employee but also costs to the employer where the two do not necessarily coincide. In other words, a fraction of the costs due to deviant behavior can be recuperated by the firm since the employee may perceives it as part of his or her compensation package, while the remainder can be considered as a “dead weight loss” (DWL).

**Assumption 4.** If the engagement with an employee is terminated, the present value of the costs of replacement is assumed to be \( R \). Initially we make here a simplifying assumption that \( R \) is an exogenous variable.
We will continue with a model with the assumptions that will be relaxed gradually. These would be advanced by the following assumptions.

**Assumption 5.** The employer has a perfect knowledge of the employee’s self-perceived quality.

**Assumption 6.** The employee is able to price the perceived quality correctly.

**Assumption 7.** Initially it is assumed that the employee benefits fully from his or her deviant behavior; i.e., the benefits of deviant behavior to the employee equal the negative value of the consequences to the firm (use of the firm’s resources). Subsequently, this assumption will be relaxed and a dead weight loss will be introduced.

Thus, the level of costs of deviant behavior during period \( t \), motivated by the difference between the level of compensation set in the beginning of the period, \( W_t \) and the desired level \( W(Q) \), is:

\[
a_t W_t = [W(Q) - \bar{W}_t] + \xi a_t W_t
\]

where \( \xi \) follows log normal distribution with a drift of \( \mu \) and standard deviation \( \sigma \),

\[
da = \mu dt + \sigma dZ.
\]

Simplifying the above, the expected level of the random component of the actual costs of deviant behavior is \( \mu a_t W_t \) where \( \mu \) is drift over the entire period \( t \). Consequently Eq. (1) becomes:

\[
a_t W_t = [W(Q) - \bar{W}_t]/(1 - \mu)
\]

The employer sets the level of compensation of the employee at the beginning of period \( t \), hoping to recuperate the consequence of the employee’s deviant behavior:

\[
\bar{W}_t = W(Q) - E[a_t W_t] = W(Q) - \mu a_t W_t
\]

where \( \mu \) reflects the risk-adjusted expected level of the random component of the costs of deviant behavior in present value terms - \( \mu \rho = \mu, \quad \rho \leq 1 \)

The “dead weight loss” is therefore:

\[
DWL_t = a_t W_t - E[a_t W_t] = [\mu - \mu \rho] a_t W_t
\]

This would equal to zero in present value terms if the employer prices the risk correctly. In other words, in this scenario the employer may never have an incentive to terminate its engagement with an employee being solely motivated by his or her deviant behavior.

Our conclusion may change if we relax one of the assumptions above:
**Assumption 5a.** (Replacing Assumption 5): The employer has no prior knowledge of the employee’s self-perceived quality and therefore of his or her costs of deviant behavior.

At time 0 when the firm sets the initial offer to a new employee,

\[ \bar{W}_0 = W(Q) - E(a_0(W_0, Q) W_0) \]

But this time the employer forms the expectations based on some average which is not specific to this particular employee:

\[ \bar{W}_0 = W(Q) - \lambda(1 + \mu*) W_0 \]

The employee reacts by choosing his or her level of deviant behavior:

\[ a_0 W_0 = [W(Q_p) - \bar{W}_0] + E[a_0 W_0] \]

Thus the initial DWL is:

\[ DWL = \frac{[W(Q_p) - W(Q)]}{1 - \mu} - \lambda(1 + \mu*) \]

If we do not assume an iterative learning process (see next scenario), or, the employer may not adjust the wage below the initial offer, then, the employer will decide to fire the employee if:

\[ \rho[DWL] > R \]

where \( \rho \) is the risk adjusted present value factor over the expected duration of the employee’s engagement.

We will relax additional assumptions. Specifically Assumption 6 and Assumption 7:

\[ a_i W_i = [W(Q_p) - \bar{W_i}] + \mu a_i W_i \]

but the employee sees only partial benefits:

\[ \bar{a}_i W_i = f\{ [W(Q_p) - \bar{W_i}] + \mu a_i W_i \} \]

For example, we may assume a simple linear function:

\[ \bar{a}_i W_i = \lambda\{ [W(Q_p) - \bar{W_i}] + \mu a_i W_i \} \]

where \( \lambda \) is a constant \( \lambda < 1 \)

The firm sets the compensation level as:
\[
\bar{W}_t = W(Q) - E(a_iW_i),
\]

and therefore

\[
DWL_t = [W(Q) - \bar{W}_i] - a_iW_i.
\]

This model, as is, does not generate a solution. Using Contract Theory, a solution can be obtained based on more restrictive assumptions. The quantitative results lead to the following hypotheses:

**Hypothesis 1.** Raising the level of compensation for an existing employee in order to diminish the level of deviant behavior, as opposed to hiring a new one, may be optimal for a very limited range over the exiting compensation level. This is true since the expected level of deviant behavior of an existing employee and that of a new employee receiving the same level of compensation, may be different due to iterative history of an existing employee.

**Hypothesis 2.** Lack of knowledge regarding the nature of newly hired employee may lead to an iterative solution, i.e. the employer may have to re-fire new employees until an optimal solution is reached.

**Hypothesis 3.** If there are no restrictions on wage setting (i.e., the level of compensation can be reduced) the employer may be better off reducing the level of compensation and let the employee decides if he or she wishes to quit, rather than actively firing the employee. If, however, there are restrictions on wage reduction, the employer may be better off actively firing the employee. The above is true given the level of certainty regarding replacement costs (as long as there is a positive likelihood that an existing employee will accept the reduction of wages and remains with the firm).

The second component approach that can be applied to the employee’s deviant behavior is a behavior model presented in the next section.

### 3.2 Prospect theory components

An attempt is made below to introduce a behavioral model in the world of Kahneman and Tversky. Based on their Prospect Theory, the employer is said to maximize a value function consisting mainly of two components: the employee’s deviant behavior damage on the one hand, and the cost of replacing the employee upon his/her dismissal on the other hand.

Assume that in Period, \( N_i \) workers are employed by a certain employer. An employee may cause damage as a result of destructive workplace deviance. The damage that the employee causes in Period is \( D_{i,t} \). If the employer dismisses the worker who caused the damage, it would incur a cost related to replacing the employee, which equals \( R_{i,t} \) dollars. The efficacy of the employer as a result of the employee’s work is positive as long as the damage he/she causes is less than the replacement costs. Thus, there are two basic ranges of damage caused by the worker’s destructive deviance.

The first range is \( D_{i,t} < R_{i,t} \). In this range the damage caused by employee in Period, \( D_{i,t} \) is less than the replacement cost, \( R_{i,t} \). The employer therefore continues to profit from the employee because the economic value of his/her performance in this case exceeds the
economic loss which results from his/her destructive deviance. The difference between profit
and loss is \((R_{i,t} - D_{i,t}) > 0\).

The second range is \(D_{i,t} > R_{i,t}\). As long as employee \(i\) in Period \(t\) causes damage greater than
his/her replacement cost, \(R_{i,t}\), it pays the employer to replace the employee. The employer loses
the difference between employee \(i\)'s damage and the replacement cost, \(R_{i,t}\). The loss then is
\((D_{i,t} - R_{i,t}) < 0\).

Let us assume that the probability \((P)\) that an employee falls within the first or the second
ranges is \(P_G\) and \(P_L\), respectively, where \(G\) and \(L\) denote gains and losses to the employer,
respectively.

**The Value Function of the Employer**

We will employ here a conceptual framework from Prospect Theory developed by Kahneman
and Tversky (1979) and Tversky and Kahneman (1992) that was also applied in numerous
studies.

The employer is supposed to have a value function \((V)\) resulting from work conducted by
the deviant employee.

When the employee belongs to the first range, the employer gains the difference \((R_G - D_G)\)
and its prospective value function \((V_G)\) equals:

\[
V_G = (R_G - D_G)^{\alpha G},
\]

where \(\alpha\), as defined by Tversky and Kahneman, is the exponent of the value function, \(G\)
denotes the gain and all other notations are as defined above. In their experimental study,
Tversky and Kahneman’s (1992), estimate for \(\alpha\) was 0.88. In this model additional estimates
for \(\alpha\) could be used as well.

When the employee belongs to the second range, the employer loses, and its value function
is negative and equals to:

\[
V_L = - \lambda (D_L - R_L)^{\alpha L},
\]

where \(L\) denotes a loss, \(\lambda\) is the loss aversion coefficient, and the other notation is as defined
above. Tversky and Kahneman’s (1992) estimate for \(\lambda\) was 2.25. This value was also adopted
by Barberis, Huang and Santos (2001). The \(\lambda\) coefficient determines how keenly losses are felt
relative to gains.

The expected value (or utility) to the employer \((E_V)\) arising from the work deviant behavior
of the two types of employees will be:

\[
E_V = W_G(P) V_G - W_L(P) V_L,
\]

where \(V_G\) and \(V_L\) are value functions for gains and losses given by Egs. (1) and (2),
respectively. \(W_G(P)\) and \(W_L(P)\) as suggested by Tversky and Kahneman (1992), are probability
weighting functions for gains and losses, respectively, and are given by:

\[
W_G(P) = P^\gamma / \left[ P^\gamma + (1 - P)^\gamma \right]^{1/\gamma},
\]

\[
W_L(P) = P^\delta / \left[ P^\delta + (1 - P)^\delta \right]^{1/\delta},
\]
where \( P \) denotes probability, \( W \) denotes weight, \( \gamma \) is the degree of risk aversion for gains, and \( \delta \) is the coefficient of risk seeking for losses. In their experimental study, Tversky and Kahneman (1992) found estimates of 0.61 and 0.69 for \( \gamma \) and \( \delta \), respectively.

To estimate the expected value for an employer, Tversky and Kahneman’s (1992) experimental estimates of the various parameters, such as \( \alpha, \lambda, \gamma \) and \( \delta \), will be employed. However, for sensitivity purposes, other values will be used as well in the context of both the simulations and the survey planned to be conducted.

4. Model boundary conditions

4.1 Responses to deviant behaviour

There are other employer responses to deviant behaviour in addition to pure toleration or dismissal. However, in this model we fix the employer responses to the extremes of either toleration or dismissal under conditions of detecting deviance that has direct monetary consequences. Thus we make a distinction between deviance resulting in direct monetary damages and less severe forms of deviance.

We realize that limiting employer responses to deviance may be unrealistic for some types of deviance, but for deviance that results in direct financial loss, we suggest that restricting employer reactions to these options is realistic. One reason is because both toleration and dismissal can be instituted immediately upon the detection of the deviant behaviour. Fast action is desirable under conditions of financial loss because such loss is likely to be manifest in greater and more obvious damage to the organization relative to milder forms of deviant behaviour. This leads to the following assumption.

**Assumption 8:** Deviance resulting in direct financial loss is likely to be manifest in greater and more obvious damage to the organization relative to milder forms of deviant behaviour.

Limiting our options toleration and dismissal is also reasonable because outcomes of deviant resulting in financial loss would warrant stronger reactions. In addition, because lesser forms of deviant behaviour are more likely to go undetected by organizational decision makers this model is more likely to be used compared to models that consider mild deviance and less severe reactions from organizations. Here we articulate two additional assumptions.

**Assumption 9:** Deviance resulting in direct financial loss is more likely to be detected relative to milder forms of deviant behaviour.

**Assumption 10:** Models involving deviance resulting in direct financial loss are more likely to be used relative to models involving milder forms of deviant behaviour.

4.2 Deviance as a calculated reaction

We recognize that there are less calculative ways for employees to express deviance and those employees are influenced by many variables that are not directly linked to calculated behaviours. For example, deviance is often presumed to result from affect and mood. However, this model is concerned with situations where deviance is a calculated reaction to financial inequity perceptions by employees.

We note that the basis of our proposed equity calculation is based on the employee’s current and desired levels of compensation, and that this is a limited perspective.
Clearly employees may react to perceived inequities in domains other than compensation, such as inequities in other forms of recognition and reward. However, for the sake of parsimony and to fit into the tradition of behavioural economics, we are assuming that employees are using financial compensation as a basis of equity calculation. This assumption is supported by abundant research in the management literature which suggests that employee’s response to pay is far more intense compared to other organizational rewards (Currall et al., 2005).

**Firm Decision Systems as Level of Analysis**

Of course, individuals are nested within firms. Our unit of analysis is essentially firm-level. However, we recognize that while responses to deviance will vary across firms such responses may also vary across units of management decision making within firms. Thus, at a finer level, our unit of analyses for the proposed models is that of different employment decision systems, whether these are nested within large firms or are measured across separate firms.

5. **Discussion**

In this paper we have proposed a behavioural model linked to behavioural economic theories. We do not attempt to explain how job attitudes or job affective states are connected to the deviant behaviours that are used as decision antecedents in this model. The behavioural nature of this model does, therefore, result in several limitations.

**Limitations**

One limitation is that the model is driven by behavioural inputs. Although researchers may wish to model other types of antecedents, such as affective and attitudinal antecedents, this would require the use of other models. A second limitation of restricting our model to behavioural dimensions is that ethical and moral considerations may be harder to build into the model. This is because the nature of intention is an important component of behavioural reasoning. For example, without being able to determine an employee’s intent when considering deviant behaviour resulting in financial loss, it may be rational to respond to deviance only by using financial information. An added implication is that the use of this model in making a decision to keep or terminate an employee from purely a rational economic point of view may be justified when it is not possible to determine an employee’s intent behind deviant behaviour.

**Theoretical Implications**

Here we note that there could be many other sources of deviance in addition to the comparison of actual versus desired levels of compensation. However, the confirmation of any of these antecedents would require qualitative methods to assess the cognitive states of individual deviant employees. By focusing simply on deviant reactions to actual versus desired levels of compensation within this proposal, our model does not specify particular antecedents to these behaviours. By implication, the use of this model disregards many potential sources of deviance which in turn limits the conceptual scope of this model.

**Practical Implications**

To consider practical implications we ask the following questions. What are the advantages of behaviourist models? And we question under what conditions would we want to use them? We suggest that when situational complexity introduces sufficient measurement barriers that a
cognitive model becomes unfeasible to implement, it may then be worthwhile to use only behavioural antecedents when modelling responses to deviance. Under our proposed model, we note that outcomes specified by the model result in severe damage and they are strongly linked to antecedent behaviours. Thus the model proposed in this paper can be useful even without affective or cognitive antecedents.

We also note that how to identify deviance could practically limit the applications of our model. Specifically, because we consider deviance resulting in financial damage in this model, and also because we consider financial damage as an outcome traced to individuals, the model will be limited to situations where such deviance can be detected and linked to an individual. Although deviance resulting in financial damage is likely to be easier to detect than other types of deviance, it may at times be difficult to link such financial losses to particular employees.

**Next Steps**

Employers should be asked about the extent of deviant behaviour among their workers. We expect that deviant behaviours resulting in financial damage have a relatively low base rate but in fact do occur. We also expect that deviant behaviours resulting in financial damage occur within cohorts of otherwise high performers. We also assume that employers are aware of losses that are directly connected to calculative deviance. However, each of these assumptions needs to be verified through future research.

We may need to develop methods to link monetary damages to the calculated deviance of a particular employee, recognizing that we must assume calculated deviance even if losses can be linked to the deviant behaviour of a specific employee. The literature on employee termination could be an area that would inform this inquiry.

Links to previous research programs should also be examined. For example, one research area that could be considered an antecedent to this model is the work on industrial sabotage in labour economics. Sabotage in tournaments is another area to examine, which would be related to industrial relations research.

**6. Conclusion**

In this paper we use two theories as complementary sources to generate our proposed relationships between deviance and employer reactions. By explicitly incorporating an economic theory – principal-agent theory – and a psychological theory – prospect theory – to account for deviant behaviour we take an interdisciplinary approach that is strength of this model. Furthermore, we are dealing with an important issue in organizational behaviour by taking an economic approach to analysis, which is unusual. Thus, while the field of investigation is organizational behaviour, our methods rest in behavioural economics. Although there are multiple areas for further inquiry that would strengthen the rationale for this model, this model presents a parsimonious means for making decisions under the stressful situation of deviant behaviours resulting in financial damage. We suggest that researcher examine the proposed next steps we develop in the preceding paragraph and call for further research using simulations or empirical data collection to validate or refine the model developed in this paper.

**Acknowledgements.** We gratefully acknowledge the help of Professors Yossi Yagil, Elli Kraizberg and Yoav Vardi in conceiving this paper.
References


