Behavioral Agency Theory

2 authors:

Alexander Pepper
The London School of Economics and Political Science

28 PUBLICATIONS   **442** CITATIONS

SEE PROFILE

Julie Gore
Birkbeck, University of London

39 PUBLICATIONS   **961** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Behavioral agency theory View project

Qualitative Research Symposium View project

Behavioral Agency Theory: New Foundations for Theorizing about Executive Compensation

Alexander Pepper

The London School of Economics and Political Science

Julie Gore

University of Surrey

Corresponding author: Alexander Pepper, Department of Management, The London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, United Kingdom

Email: a.a.pepper@lse.ac.uk

ABSTRACT

This paper describes new micro-foundations for theorizing about executive compensation, drawing on the behavioral economics literature and based on a more realistic set of behavioral assumptions than those which have typically been made by agency theorists. We call these micro-foundations "behavioral agency theory". In contrast to the standard agency framework, which focuses on monitoring costs and incentive alignment, behavioral agency theory places agent performance at the center of the agency model, arguing that the interests of shareholders and their agents are most likely to be aligned if executives are motivated to perform to the best of their abilities. We develop a line of argument first advanced by Wiseman and Gomez-Mejia (1998), and put the case for a more general reassessment of the behavioral assumptions underpinning agency theory. A model of economic man predicated on bounded rationality is proposed, adopting Wiseman and Gomez-Meija's assumptions about risk preferences, but incorporating new assumptions about time discounting, inequity aversion and the trade-off between intrinsic and extrinsic motivation. We argue that behavioral agency theory provides a better framework for theorizing about executive compensation, an enhanced theory of agent behavior and an improved platform for making recommendations about the design of executive compensation plans.

INTRODUCTION

Agency theory has been a major component of the economic theory of the firm since the publication of formative work by Spence and Zeckhauser (1971), Alchian and Demsetz (1972), Ross (1973) and Jensen and Meckling (1976). It has also become the dominant theoretical framework for academic research on executive compensation (Bratton, 2005). The literature on senior executive reward is now very extensive, drawing on a variety of scholarly traditions, including economics, law, organization studies, accounting and finance. In addition to the agency approach, theoretical frameworks include tournament theory (Lazear & Rosen, 1981), human capital theory (Combs & Skills, 2003), the managerial-power hypothesis (Bebchuk, Fried & Walker, 2002), institutional theory (Balkin, 2008), political theories (e.g., Ungson & Steers, 1984) and theories about fairness (e.g., Wade, O'Reilly & Pollock, 2006). Literature reviews and summaries are provide by Gomez, Meija & Wiseman (1997), Devers, Cannella, Reilly and Yoder (2007) and Gomez-Mejia, Berrone and Franco-Santos (2010: 117-140). Denvers et al (2007) note that behavioral research is a relative new feature of the literature on senior executive reward.

That agency theory has shortcomings has been apparent for some time. Most notably, given Jensen's role as a leading agency theorist, empirical work carried out by Jensen and Murphy (1990) failed to establish a conclusive link between CEO pay and stock price performance[1] Ten years later, in a meta-analysis of 137 empirical studies, Tosi, Werner, Katz and Gomez-Mejia (2000) similarly found that incentive alignment as an explanatory agency construct for CEO pay was at best weakly supported by the evidence. More recently, Frydman and Jenter (2010) have argued, based on a review of US executive compensation data covering the period 1936 to 2005, that neither optimal contracting (agency theory) nor the managerial power hypothesis is fully consistent with the available evidence. Roberts, another agency theorist, has commented that agency theory performed poorly during the

financial crisis and has reported various situations where strong incentives are evidently not optimal, as agency theory implies (Roberts, 2010). These include when good measures of an agent's effort or performance are not available, when multi-tasking is required, and when cooperation between different agents is necessary, all common situations where top management teams are concerned. Roberts puts forward arguments in favor of implementing *weak* rather than *strong* incentives in such circumstances.

This paper proposes a new version of agency theory which provides a better explanation of the connection between executive compensation, agent performance, firm performance and the interests of shareholders. We call this "behavioral agency theory", developing a line of argument first advanced by Wiseman and Gomez-Mejia (1998), who proposed that the normal risk assumptions of agency theory should be varied to incorporate ideas from prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Sanders and Carpenter (2003) have subsequently adopted a behavioral agency perspective in their examination of stock repurchase programs and a summary of the literature using the behavioral agency framework is provided by Finkelstein, Hambrick and Cannella (2009). Rebitzer and Taylor (2011) provide a general examination of behavioral approaches to agency and labor markets in the 4th edition of Ashenfelter and Card's influential handbook on labor economics. However, a settled theory and agreed terminology for the behavioral agency model does not yet exist. In contrast to the standard agency framework, which focuses on monitoring costs and incentive alignment, behavioral agency theory places agent performance and work motivation at the center of the agency model, arguing that the interests of shareholders and their agents are most likely to be aligned if executives are motivated to perform to the best of their abilities, given the available opportunities. Behavioral agency theory builds on four constructs which have been identified as key factors affecting behavior by behavioral economists (Camerer, Loewenstein & Rabin, 2004). These are: (1) loss aversion and

3

reference dependence; (2) preferences relating to risky and uncertain outcomes; (3) temporal discounting; and (4) fairness and inequity aversion. It incorporate a theory (crowding out) relating to the trade-off between intrinsic and extrinsic motivation (Frey & Jegen, 2001; Sliwka, 2007). It also introduces goal-setting theory (Locke & Latham, 1984, 1990) to the agency model, on the basis that it represents a pragmatic way of contracting between principal and agent.

The paper proceeds as follows: it begins by describing agency theory's main elements and underlying assumptions, before reviewing the limitations of positive agency theory as an explanation of the relationship between senior executives and shareholders, and reconceptualizing what is meant by economic man (i.e., homo economicus of neoclassical economics). It continues with an explanation of the behavioral agency model, describing the main component systems and commenting in some detail on the significance of motivation, risk, time discounting, inequity aversion and goal setting. It examines the relationship between job performance and firm performance, discusses ways in which behavioral agency theory departs from standard agency theory, and considers the implications of behavioral agency theory for compensation design, before concluding.

POSITIVE AGENCY THEORY

Positive agency theory[2], the standard model of agency which we consider in this paper, has been extensively used as a basis for theoretical and empirical work by management scholars and organization theorists (e.g., Eisenhardt, 1989), as well as being widely applied in examining research questions relating to executive compensation (e.g., Tosi & Gomez-Mejia, 1989). It argues that the firm is a special case of the theory of agency, that a firm provides a nexus for a complex set of contracts, both written and unwritten, between various parties, and that agency costs are generated as a result of the different interests and contractual

arrangements between owners and top managers (Alchian & Demsetz, 1972; Jensen, 1983; Jensen & Meckling, 1976). The underlying assumptions are that organizations are profit seeking, that agents are both rational and rent seeking, and that there is no non-pecuniary agent motivation[3]. It is further assumed that principals are risk neutral, because they can balance their portfolios, that agents are risk averse, because the potential wealth effects of the employment relationship are significant, that an agent's utility is positively contingent on pecuniary incentives and negatively contingent on effort, and that time preferences are calculated mathematically according to an exponential discount function (Jensen, 1998). It is postulated that effort and motivation increase monotonically with additional reward[4]. The pay-effort function is visualized as a straight line with a positive gradient proceeding from bottom left to top right.

Efficiency is the main criterion for assessing the success or otherwise of programs under agency theory. Agency theory focuses on the costs of the potential conflict of interest between principals and agents, referred to as "agency costs". Jensen and Meckling define agency costs as the sum of the monitoring expenditures of the principal, the bonding expenditures of the agent, and the residual loss in welfare experienced by the principal as a result of the divergence of interests between the principal and the agent (Jensen & Meckling, 1976). Jensen subsequently offers a broader definition, describing agency costs as "the sum of the costs of structuring, bonding and monitoring contracts between agents…[which]…also include the costs stemming from the fact that it does not pay to enforce all contracts perfectly" (Jensen, 1983: 331)[5]. Agency costs are thus a special case of transaction costs (in a Coasian sense) in their internal (intra-firm) rather than external (intra-market) form. Positive agency theory proposes that principals can mitigate agency costs by establishing appropriate incentive contracts and by incurring monitoring costs. This is formalized by Eisenhardt in two propositions - first, in respective of incentives: "when the contract between

the principal and agent is outcome based, the agent is more likely to behave in the interests of the principal" (Eisenhardt, 1989: 60); secondly, in respect of monitoring: "when the principal has information to verify agent behavior, the agent is more likely to behave in the interests of the principals" (Eisenhardt, 1989: 60).

BEHAVIORAL AGENCY THEORY

Behavioral agency theory argues that the model of economic man which forms the micro-foundations of agency theory is too simplistic. It proposes a reconceptualization, developing a new model which assumes bounded rationality[6], recognizes the importance of agents' human capital (taking this to be a function of ability and work motivation) and allows for departures from the rational choice model when it comes to loss, risk and uncertainty aversion, time discounting, inequity aversion and the trade-off between intrinsic and extrinsic motivation. It proposes that the standard agency theory model of man should be modified in a number of ways. The first modification relates to agent performance and work motivation. Agency theory places less emphasis on the objective of motivating agents to perform to the best of their ability than it does on aligning the interests of agents and principals. Leibenstein (1966) argues that, given the importance of what is now called human capital, motivation (in particular, intrinsic motivation) cannot be ignored in the economic calculus. Pratt and Zeckhauser (1985) make the same case for agency theory. Behavioral agency theory argues that maximizing agent performance should be a key objective of the principal-agent relationship and that the importance of the agent's work motivation, including intrinsic motivation, should not be underestimated. It challenges the idea that intrinsic and extrinsic motivation are either independent or additive, arguing instead that contingent monetary rewards might actually cause a reduction in intrinsic motivation (see Deci & Ryan, 1985).

Frey and Jegen (2001), following a line of scholarly thinking that dates back to Lepper and Greene (1978), have described this phenomenon as "crowding-out"(see also Sliwka, 2007) .

The second modification relates to risk and uncertainty[7]. Behavioral agency theory assumes that senior executives are primarily loss averse and only secondarily risk averse (Wiseman & Gomez-Mejia, 1998). Gains and losses are calculated by each individual agent in relation to a reference point which he or she subjectively determines. Risk preferences differ in gains and losses, resulting in an "S-shaped" value function, with losses looming larger than gains. This means that, below a reference point, agents will be loss averse, resulting in an increase in his or her appetite to take short term risk. Above the reference point agents will generally be risk averse, but decision weights will vary depending on subjective probability assessment; for example, small probabilities are over-weighted and large probabilities are under-weighted.

The third modification relates to time preferences. In behavioral agency theory it is assumed that agents discount time according to a hyperbolic discount function, rather than exponentially, as is the case with financial discounting (Ainslie, 1991; Ainslie & Haslam, 1992). This means that future rewards are heavily discounted and allows for the possibility of preference reversals. Actual average discount rates vary between individuals and must be determined empirically.

The fourth modification relates to an agent's perceptions of equitable compensation. If agents feel that their inputs, the effort and skills which they put in to their work, are fairly and adequately rewarded by outputs, the tangible and intangible rewards from employment, then the agents will be happy in their work and motivated to continue to contribute at the same or at a higher level (Adams, 1965). However, if the relationship between inputs and outputs is not proportionate, then an agent will become dissatisfied and hence demotivated. In this model the agent's equity benchmark is subjectively determined according to market norms

and personal referents. Fehr and Schmidt (1999: 819) call this phenomenon "inequity aversion". As is the case with risk and time discounting, we anticipate that actual levels of inequity aversion will vary between individuals and must be determined empirically.

Table 1 summarizes the assumptions about the characteristics of economic man which provide the foundations of agency theory and compares them with the way in which behavioral agency theory reconceptualizes the model. An important early conclusion which can be drawn is that an agent's perception of the (subjectively-calculated) value of an incentive award will typically be less than the award's (objectively-calculated) economic value. This clearly has implications for the way that incentive contracts are designed.

```
--------------------------------------------------
|                                                |
|              TABLE 1 ABOUT HERE                |
|                                                |
--------------------------------------------------
```

Assessment Criteria and Unit of Analysis

Behavioral agency theory proposes that it is necessary to use both effectiveness and efficiency as yardsticks for judging agent activity. By adopting effectiveness as well as efficiency as criteria for assessment we follow a long line of management theorists dating back to Barnard (1938 |1968)[8]. Simon (1945 |1997) pointed out that the terms "effectiveness" and "efficiency" were considered to be almost synonymous until the end of the 19th century and were generally thought to mean the power to accomplish the purpose intended; however, the meanings of the two words subsequently diverged. Efficiency came to be defined, firstly in engineering and subsequently in economics, business, and management, in terms of the relationship between inputs and outputs. In this paper we use the terms efficiency and effectiveness in the following way: on the one hand, an action, event, plan, policy or program is considered to be *efficient* if it causes inputs to be minimized for a

given set of outputs or outputs maximized for a given set of inputs; on the other hand, an action, event, plan, policy or program is considered to be *effective* if it is capable of achieving its intended objectives, what the objectives are being exogenous to the theory. We contend that it is necessary for management scholars to adopt both criteria in order to provide a complete and accurate evaluation of management policies, plans and programs. Taking executive compensation as an example, a compensation plan might be effective and efficient (i.e., achieve its objectives of motivating top managers and aligning the interests of managers and shareholders, doing so in such a way that costs are minimized), effective but not efficient (i.e., achieve its objectives but in a way that is more costly than necessary), or neither effective nor efficient (i.e., fail to achieve its objectives at the same time as being costly). However, we argue that it makes no sense to describe a management plan or program as efficient but not effective. The concept of effectiveness is already implied by the concept of efficiency; a lower cost (or indeed no cost at all) could otherwise be incurred while still failing to achieve the desired objectives.

An important premise of behavioral agency theory, consistent with the top management team or "upper echelons" approach (Hambrick & Mason, 1984) is that senior executive teams have a major impact on firm performance. We define "top management team" (and hence "top manager") as the group of very senior executives who are responsible for defining and executing a firm's strategy, who through their actions are capable of affecting the company's profits, share price, reputation and market positioning (Carpenter, Geletkanycz & Sanders, 2004). This group, which includes the chief executive officer (CEO), the chief operating officer (COO), the chief financial officer (CFO), divisional heads and other heads of function, is sometimes referred to as the "management board", "operating board ", "executive committee" or "general management committee". Changing trends in corporate governance mean that, while historically these individuals would have been executive directors, it is

increasingly common in many countries to find only the CEO and CFO on the main board, while all the key senior executives sit on the executive committee, or equivalent (Pepper, 2006). By defining top managers in this way, this part of behavioral agency theory becomes, in a sense, tautological (corporate performance is, in part, a function of the performance of top managers; top managers are those individual agents who are able to influence corporate performance). However, this is the type of "useful tautology" which Jensen (1983: 330-331) points out is a necessary part of the process of theory development; nor does its inclusion in behavioral agency theory mean that this part of the theory becomes in practice irrefutable - it might be demonstrated in certain cases that top managers are not in practice able to have a significant impact on firm performance.

Unlike upper echelons theory, which takes the top management team as the primary unit of analysis (Hambrick & Mason, 1984) behavioral agency theory focuses on the behaviors, interests and actions of individual top managers or agents. Following Boxall and Purcell (2003) we model an agent's performance as a manager of a large firm as a function of his or her ability, motivation and opportunity. Agents will perform if they have the ability (the necessary knowledge, skill and aptitude), the motivation (intrinsic and extrinsic), and the right opportunities (including the necessary work structures and business environment); formally:

$$P_{\bar{a}} = f(A, M, O) \qquad (1)$$

where $P_{\bar{a}}$ stands for the job performance of the agent, A stands for ability, M stands for motivation or "motivational force", after Lewin (1938), and O stands for the agent's opportunity set.

Boxall and Purcell conceptualize ability in much the same way that Becker (1993) conceptualizes human capital, i.e., in terms of knowledge, skills, health, value and habits. Leibenstein (1966) comments on the importance of motivation to human capital. The

significance for behavioral agency theory is that a competent agent must be properly motivated in order to ensure optimal performance (Pratt & Zeckhauser, 1985), meaning in this context the point where efficiency is maximized subject to any effectiveness constraints, and effectiveness is maximized subject to any efficiency constraints. Thus we define human capital in this article as "motivated ability" rather than merely as a function of education and experience.

In this paper we focus on the role of motivation in influencing the job performance of agents. For the purposes of the current paper, we take ability, which has its roots in the learning and development and human capital literatures, and opportunity, which can be traced to the leadership and strategy literatures, as given.

Motivation

The theory of work motivation most commonly used in investigations into the motivational impact of pecuniary incentives is expectancy theory (Vroom, 1964). According to expectancy theory, motivational force is a function of expectancy (the strength of belief or subjective probability that an action i will lead to a particular outcome j), instrumentality (the degree to which a first outcome j will lead to a second outcome k), and valence (the preference which an individual has for the second outcome k)[9]. Expectancy theory is essentially concerned with extrinsic, rather than intrinsic or total motivation. Thus expectancy theory can be formally represented as:

$$X_i = f\left(E_{ij}, I_{jk}, V_k\right) \tag{2}$$

where X_i is the extrinsic motivational force to perform act i, E_{ij} is the strength of expectancy that act i will be followed by outcome j, I_{jk} is the instrumentality of outcome j for attaining outcome k, and V_k is the valence of outcome k. Expectancy theory thus describes a cognitive process and is distinct from many of the other standard theories of motivation, especially

theories based on needs, drives and learned behaviors, which seek to explain the psychological content of motivation

Steel and König (2006) have proposed a modified version of expectancy theory which they call "temporal motivation theory". It postulates that motivation can be understood in terms of valence and expectancy[10], weakened by delay, influenced by risk and uncertainty, with different valences for gains and perceived losses. Temporal motivation theory brings expectancy theory together with prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) and hyperbolic discounting (Ainslie, 1991; Ainslie & Haslam, 1992). Reducing Steel and König's formula down to its minimal form gives:

$$X_i = \left\{ \frac{E_{ik}^{pt} \times V_k^{pt}}{1 + \delta t} \right\} \tag{3}$$

where X_i is again the extrinsic motivational force to perform act i, E_{ik}^{pt} is the expectancy that act i will lead, via j, to outcome k, V_k^{pt} is the valence for outcome k, δ is the personal discount factor for the delay between act i and outcome k, and t represents the time-lag. Expectancy and valence are both calculated in accordance with prospect theory. The main implications of this are that probabilities and decision weights are determined subjectively and valence is affected by risk perception: in particular, valences will differ significantly depending on whether gains or losses are expected[11]. Time effects are determined by a hyperbolic discount function after Ainslie (1991) rather than the more conventional exponential discounting function used in financial theory. This means that, in Steel and König's revised expectancy model, the valence which an agent attaches to k takes into account risk and uncertainty, as well as being discounted for any time delay between the occurrence of act i and outcome k.

Positive agency theory places less emphasis on the objective of motivating agents than it does on alignment of the interests of agents and principals. Deci and Ryan (1985) point out that there are two distinct forms of motivation, intrinsic motivation, where an agent performs

12

an activity for its inherent satisfaction rather than because of some separable consequence, and extrinsic motivation, where an activity is carried out because of its instrumental value. Kreps (1997) argues that it is not necessary to postulate the concept of intrinsic motivation on the basis that what is called intrinsic motivation may in fact be no more than a series of vaguely defined extrinsic motivators. Besley and Ghatak (2005) contend, on the contrary, that there is such a thing as a motivated agent whose economic behavior is affected by intrinsic motivation, but their argument is directed towards employees of public sector or non-profit organizations whose activities coalesce around a "mission". Deci and Ryan (1985) argue that the importance of intrinsic motivation should not be underestimated. They challenge the idea that intrinsic and extrinsic motivation are either independent or additive, arguing instead that contingent monetary awards might actually cause a reduction in intrinsic motivation. Boivie, Lange, McDonald and Westphal (2011) point out how, in the case of CEOs, high organizational identification, which may be associated with intrinsic motivation, can help to reduce agency costs. Frey and Jegen (2001) and Sliwka (2007) postulate that in some cases extrinsic rewards can "crowd-out" intrinsic motivation, particularly if monetary incentives are badly designed. They argue for a strong form of crowding-out whereby an increase in extrinsic reward leads to an overall reduction in total motivation. A weaker form of crowding-out, whereby the level of total motivation is maintained only if the increase in extrinsic reward more than compensates for the reduction in intrinsic motivation, can alternatively be postulated.

Following Deci and Ryan, (1985), the relationship between intrinsic and extrinsic motivation can be stated formally as follows:

$$M_i = f(N_i, X_i) \tag{4}$$

where M_i is an agent's total motivational force, N_i is the agent's intrinsic motivation, and X_i is the agent's extrinsic motivation. M_i, N_i, and X_i can be thought of in terms of stimuli, actions

or outcomes i.e., M_i represents motivation resulting from i, where i is a stimulus or bundle of stimuli, an action or package of actions, an outcome or collection of outcomes. However, the relationship between N_i and X_i is neither linear nor additive (Deci & Ryan, 1985). In a dynamic sense, when changes in incentives occur, there is evidently a trade-off of some kind between the two types of motivation. Whether this is more accurately described by the strong crowding-out conjecture, where a change in extrinsic motivation as earnings increase from e to g, $(+\Delta X_{eg})$ leads to a decrease in intrinsic motivation $(-\Delta N_{eg})$ such that $\Delta N_{eg} > \Delta X_{eg}$ and $M_e > M_g$, or by the weak crowding-out conjecture, such that $\Delta N_{eg} = \Delta X_{eg}$ and $M_e = M_g$, can only be determined empirically. This argument leads to the first two research propositions:

> *Proposition 1a: (The weak crowding-out conjecture) Above a certain level of compensation (represented by inflection point λ_1 on the agent's pay-effort curve) intrinsic motivation will decrease as compensation increases, such that the rate of increase of total motivation will diminish and will eventually, at a higher level of compensation (represented by inflection point β on the agent's pay-effort curve), reach zero.*

> *Proposition 1b: (The strong crowding-out conjecture) If compensation continues to increase above the higher level of compensation represented by inflection point β on the agent's pay-effort curve, then total motivation will start to decline as intrinsic motivation is crowded out by extrinsic rewards.*

Compensation comprises the sum of all incentives and rewards, pecuniary and non-pecuniary, arising from the agency relationship. The difference between incentives and rewards is that incentives are determined ex ante (i.e., prior to performance, thus encouraging agents to act in a particular way) whereas rewards are determined ex post.

Risk

A standard assumption of agency theory is that agents are risk averse. According to behavioral agency theory this is an oversimplification. We argue that extrinsic motivation and agent behavior are significantly affected by the agent's risk profile and that a more sophisticated model of risk and uncertainty is accordingly required. Behavioral agency theory postulates, after Wiseman and Gomez-Meija (1998), who in turn cite Kahneman and Tversky (1979) and Tversky and Kahneman (1992), that agents are primarily loss averse and consequently, contrary to one of the standard assumptions of agency theory, may actually have a high propensity to take short term risks below a certain level of compensation representing the point where perceived gains become perceived losses. Above this gain/loss inflection point, agents will generally be risk averse, but small probabilities are typically over-weighted and large probabilities are typically under-weighted. The gain/loss inflection point is itself context dependent and a matter of individual differences. In particular it is contingent upon the agent's perception of his or her individual compensation endowment which comprises their actual current compensation, enhanced to the extent of future incentives which are expected to be received with a reasonable degree of certainty. For example, a future bonus which is guaranteed or otherwise strongly anticipated based, say, on the pattern of past bonus payments, would be taken into account in the current compensation endowment, albeit discounted for future payment. In a similar way, an agent with underwater options (where the current stock price is below the option strike price) may regard this as representing a loss on his or her current compensation endowment.

This enables us to advance, following Wiseman and Gomez-Meija (1998), two further propositions:

> *Proposition 2a: Below a level of compensation (represented by inflection point λ_2 on the agent's pay-effort curve) agents are loss averse.*

Proposition 2b: Above a level of compensation (represented by inflection point λ_2 on the agent's pay-effort curve) agents are risk averse.

Time Discounting

Positive agency theory assumes that time differences can be accounted for by the type of conventional exponential discount function used in finance theory. However, behavioral economists have identified a series of anomalies in the way that individuals account for time, including preference reversal and weakness of will (undertaking actions which in the short term are pleasurable, but which agents know to be detrimental to their well-being in the long term). Ainslie (1991) explains these anomalies by arguing, based on experimental evidence, that his subjects discount future events hyperbolically so that the implied discount rate varies over time, rather than exponentially, which would require a constant discount rate. That economic agents typically discount time hyperbolically is generally accepted as the norm by behavioral economists (Frederick, Loewenstein & O'Donoghue, 2002; Graves & Ringuest, 2012). Steel and König (2006) argue that expectancy theory must take into account time differences as compensation (outcome k in Equation 3 above) may not be received until sometime after the action which leads to the payment (act i). They also argue that time differences should be accounted for using a hyperbolic discount function. Accordingly, we postulate that an agent's extrinsic motivation is affected by time discounting, calculated on a hyperbolic discount basis, as set out in the next proposition:

Proposition 3: Agents discount future compensation according to a hyperbolic discount factor such that the average discount rate δ is significantly greater than the equivalent financial discount rate.

Inequity Aversion

Behavioral agency theory postulates that motivational force is affected by inequity aversion, based on equity theory (Adams, 1965). It is widely recognized that an individual's satisfaction with his or her compensation depends not just upon buying-power, but also on how their compensation compares with the total rewards of salient others (Shafir, Diamond & Tversky, 1997). Akerlof (1982) postulates the fair-wage hypothesis according to which workers have a conception of a "fair-wage" such that, if actual earnings are less than the fair-wage, then only a corresponding fraction of normal effort will be supplied. According to Adams (1965) people seek a fair balance between what they put into their jobs (including energy, commitment, intelligence and skill – collectively "inputs") and what they get out (including financial rewards, recognition, and opportunities for personal growth – collectively "outputs"). Agents form perceptions of what constitutes an appropriate balance between inputs and outputs by comparing their own situations with those of other people in accordance with the ratio $\{O_{\bar{a}}/I_{\bar{a}}\} : \{O_r/I_r\}$ (which we refer to below as the "Adams' ratio") where $O_{\bar{a}}$ is the agent's outputs e.g., their compensation, $I_{\bar{a}}$ is the agent's inputs e.g., their skills and effort, O_r is the outputs of the agent's referents and I_r is the referents' inputs. Referents may be internal (peers, immediate subordinates, immediate superiors) or external (people doing equivalent jobs in other organizations). If agents feel that their inputs are fairly and adequately rewarded by outputs, the equity benchmark being subjectively perceived from market norms and other reference points, then they will be happy in their work and motivated to keep contributing at the same or a higher level. However, if the relationship between inputs and outputs is not proportionate, such that $\{O_{\bar{a}}/I_{\bar{a}}\} < \{O_r/I_r\}$, then the agent will become dissatisfied and hence demotivated. "Inequity aversion", as Fehr and Schmidt (1999: 819) call this phenomenon, is translated by Michelman into economic terms as "demoralization costs" (Michelman, 1967: 1214). Gomez-Mejia and Wiseman (1997) argue

that inequity aversion applies equally to senior executives as to other workers. Accordingly, we generate the next proposition:

Proposition 4: Individual agents will determine a level of compensation (represented by inflection point λ₃ on the agent's pay-effort curve) by reference to the compensation of a class of significant referents, such that the agent will tend to be dissatisfied and hence demotivated if his or her actual earnings are less than λ₃.

It is important to note for Proposition 4 that an individual agent's assessment of relative compensation levels will take account of his or her perception of their contribution in comparison with that of his or her referents, in accordance with what we have described above as the "Adams' ratio", according to which individuals seek to balance perceived relative inputs and outputs (Adams, 1965).

Goal Setting, Contracting and Monitoring

We turn now to goal setting, contracting and monitoring. We argue that these activities should be seen as integral to behavioral agency theory: goal setting and monitoring are important factors in legal contracting, which is a key element in the relationship between principal and agent (Grossman & Hart, 1983; Hart, 1995); they have also been demonstrated to be an important component of agent motivation (Locke & Latham, 1984, 1990). Goal setting theory postulates a strong connection between goals, commitment and performance. Goals must be specific, difficult, attainable, and self-set or explicitly agreed to for the motivational affect to be maximized. Much of the empirical work supporting goal setting theory has been carried out in an industrial context (e.g., with loggers, truck drivers and word processing operators). Nevertheless, behavioral agency theory postulates that many of the features of goal setting theory are generalizable to senior executives. Locke and Latham (2002) make three points which are particularly pertinent to agency relationships. First, they

argue that monetary incentives enhance goal commitment but have no substantive effect on motivation unless linked to goal setting and achievement. Secondly, they explain, through a model which they call the "high performance cycle", how goal setting and achievement together lead to high performance, in turn leading to rewards, high job-satisfaction and self-efficacy. Thirdly, they suggest a possible connection with prospect theory, both theories stressing the importance of reference points in cognition.

One of the main problems with the relationship between principals and agents which has been identified by agency theorists is that agency contracts are inevitably incomplete (Grossman & Hart, 1983; Hart, 1995). If principals were able to specify completely all that they required of their agents, then there might be no need for incentive contracts to align the interests of principals and agents - monitoring of actions and outcomes might suffice. However, in practice there are limits on knowledge and cognition. One of the reasons that principals employ agents is for the agents' expertise. An agent who is more knowledgeable about the matters which are to be specified in a contract may be able to second-guess the principal during and after contract negotiation to the agent's advantage and the principal's detriment. There are also dynamic constraints. Over the course of time the business environment which provides the backdrop for the agency contract inevitably changes. Actions which are contractually required of the agent when a contract is negotiated may cease to be appropriate at a later date because of environmental changes, and other actions which could not have been anticipated ex ante may subsequently become necessary ex post. It is contractual uncertainties of this kind that Roberts (2010) is referring to when he advocates the merits of weak rather than strong incentives in agency relationships. Goal setting, especially when it involves discussions between principal and agent about the appropriate level of objectives, is a pragmatic way of contracting, given limits on knowledge and cognition. It is also a signaling mechanism, indicating to one of the parties in an exchange relationship, the

agent, what is required by another party, the principal. Spence (1973) has shown how signaling mechanisms of this kind form an important part of economic exchange in the context of employment. Thus, goal setting, monitoring and reward, as part of a regular high performance management cycle, provide a way of improving the quality of contracting in a manner which helps to enhance rather than undermine agent motivation. This leads to a further proposition:

> *Proposition 5a: The existence in a firm of a system of goal setting, monitoring, and linked rewards and incentives for agents who are members of the top management team is positively correlated with agent performance and work motivation.*

Some care is required, however. First, it would not possible to specify in a performance contract a full set of the objectives which would be necessary to cover all possible situations that might arise during the course of a performance cycle. According to the principle of requisite variety, a control system requires a response mechanism for every exogenous shock which it might face (Ross Ashby, 1956 |1976). Top managers face great complexity in their work and it would not be possible to anticipate every possible exogenous shock in a performance contract, nor to specify fully all the requirements of the job (Mintzberg, 1997, 2009). Arrow (1985) notes how unrealistic such a complex fee function would be. Secondly, the knowledge constraints of the bounded rationality assumption place cognitive limits on an agent's ability to assimilate and understand complex goals and performance criteria. This in turn leads to Proposition 5b, which is consistent with the conclusions reached by Roberts (2010), described in the introduction:

> *Proposition 5b: Weak incentives are a more effective and efficient way of motivating agents than strong incentives.*

Agent's Job Performance and Work Motivation Cycle

The various elements of the subsystem which models agent job performance and work motivation are summarized in Figure 1. This figure illustrates the trade-off between intrinsic and extrinsic motivation (the subject of Propositions 1a and 1b), the roles played by risk (Propositions 2a and 2b), time discounting (Proposition 3) and inequity aversion (Proposition 4). The goal setting, contracting and monitoring process (Propositions 5a and 5b) are illustrated, along with the integral feedback mechanism. Two further propositions, developed later in the paper, are also represented.

INSERT FIGURE 1 ABOUT HERE

In our analysis of the trade-off between intrinsic and extrinsic motivation, risk, and inequity aversion we have identified three compensation inflection points on the agent's pay-effort curve: λ_1 which is critical to the trade-off between intrinsic and extrinsic motivation; λ_2 which determines where an individual's risk appetite changes from loss aversion to risk aversion; and λ_3 which acts as the reference point for comparisons with salient others in the context of inequity aversion. As we have explained, λ_1, λ_2 and λ_3 are critical points in the various sub-systems. In our representation of an agent's pay-effort function in Figure 2 we make the assumption that these three inflection points are identical for any one individual agent. There is support for this assumption in the argument advanced by Deci and Ryan (1985) that the psychological sub-systems for intrinsic motivation, risk, and inequity aversion converge upon a common psychological state in which cognitive, affective and conative

variables are optimally aligned. However, we assume the equality of the three inflection

points largely for mathematical convenience. In practice, even if there is a linear range

between an upper inflection point (λ_1) and a lower inflection point (λ_2 and λ_3), or a plane with

three separate inflection points, the main argument, which is that there is a set of values for

which an agent's pay-effort ratio is optimized, would not be undermined - the range of

possible outcomes would simply be expanded.

Intrinsic motivation is represented in Figure 2 by the $\varepsilon = f(N_i)$ curve and extrinsic

motivation by the $\varepsilon = f(X_i)$ curve. By superimposing the extrinsic motivation curve on top of

the intrinsic motivation curve, we generate the total motivation or $\varepsilon = f(N_i, X_i)$ curve.   This

runs parallel to the extrinsic motivation curve until total compensation reaches ω^*, at which

point crowding out sets in, intrinsic motivation starts to decline and the rate of increase of

total motivation slows accordingly.

```
┌─────────────────────────────────────────┐
│                                          │
│            FIGURE 2 ABOUT HERE           │
│                                          │
└─────────────────────────────────────────┘
```

By assuming the equality of λ_1, λ_2 and λ_3 (i.e., λ) we infer that there is a preferred level of

pay at which point the relationship between agent motivation (ε^*) and total compensation

(ω^*) is optimized, subject to constraints for risk, time discounting and inequity aversion.

This is the point when an agent's effort-to-pay ratio is maximized, such that the gradient of

the total motivation curve is equal to 1.   It implies that there is a set of first best

compensation strategies, being combinations of fixed and variable pay, contingent and

discretionary bonuses, and short-term and long term incentives: formally, that λ is represented

by the set $\{\sigma_1, \sigma_2,\dots\sigma_n\}$ where σ represents a compensation mix with a unique combination of

fixed, variable, contingent, discretionary, current and deferred rewards. If in practice there was a linear range between an upper and lower inflection point, or a plane with three separate inflection points, then this would simply increase the set of first best pay combinations. Based on this analysis we advance the next proposition:

> *Proposition 6: There is a set of first best compensation strategies combining fixed and variable pay, contingent and discretionary bonuses, and short-term and long term incentives, such that the relationship between pay and agent motivation is optimized.*

Figure 2 also illustrates a number of other phenomena: below point α motivation falls away rapidly as a result of inequity aversion – effort levels are only restored at point α when the Adams' ratio recovers to an acceptable level; above point β crowding-out means that intrinsic motivation has more or less been eliminated entirely and total motivation has peaked. If the strong crowding-out conjecture is correct, then at point γ the intrinsic motivation curve moves from positive to negative and total motivation begins to decline.

It is important to understand what this figure does and does not tell us about executive compensation. In Figure 2, λ represents the point where total compensation, comprising fixed pay, incentives and rewards, is at its most efficient and effective, and an agent's effort- to-pay ratio is at its highest. The actual pay of senior executives, which is in practice influenced by other factors such as (often imperfect) labor market conditions, strategic (inter-firm) rivalry, and political (intra-firm) gaming, may in practice be higher. Executives might be prepared to offer more effort for more incentive pay, but the marginal cost to the employing company of increasing incentive payments may be very high. This is consistent with the phenomenon of high executive compensation (which may be effective but is not necessarily efficient) and also with proposition 5b, that weak incentives are a more effective and efficient way of motivating agents than strong incentives.

Corporate Performance

A complete theory of agency must explain the mechanism which links the job performance of an executive with the performance of the firm. We take as starting point upper echelons theory (Carpenter *et al.*, 2004; Finkelstein *et al.*, 2009; Hambrick & Mason, 1984) which postulates a causal connection between business performance (the dependent variable), the cognitive skills of top managers, their observable personal characteristics (e.g., age, education, experience, socioeconomic background etc.), their strategic choices, and the objective situation (independent variables). We first simplify this a little by taking corporate financial performance to be a function of an agent's performance (as described in the motivation cycle), the performance of other agents, and the external business environment. We then build on the upper echelons approach by postulating a link between the performance of an individual agent $\bar{a}$ (itself a function of his or her ability, motivation and opportunity set), the performance of other agents $\bar{o}$ who, together with agent $\bar{a}$, comprise the top management team, the business strategy (as devised and implemented by the top management team) and the business environment, on the one hand, and business performance on the other hand.

The external business environment is largely outside the control of senior management and hence exogenous to behavioral agency theory. The job performance of other agents, $P_{\bar{o}}$, is endogenous. Indeed, the motivation and performance cycles described in this paper are replicated for all agents fitting the definition of top managers. This generates a final research proposition, that incentive compatibility between agents is a necessary condition of optimal corporate performance. We articulate this as follows:

> *Proposition 7: The incentives and rewards of individual agents must be compatible with the incentives and rewards of other agents working as part of the same top management team if firm performance is to be optimized.*

It means, for example, that agents' goals and performance conditions attaching to incentives must be compatible one with another. It also requires inequity aversion to be taken into account within the top management team - there is a strong presumption that individual agents will regard other agents in the same top management team as among their pool of referents for the purposes of equity theory. The desirability of compatible incentives is consistent with the argument that interventions may be necessary in order to align the interest of different members of top management groups (Carpenter *et al.*, 2004; Hambrick, 1994). The incentive compatibility proposition also provides a further argument in favor of weak rather than strong incentives. Roberts (2010) notes that strong incentives may not be appropriate when cooperation between different agents in necessary. Teece, Pisano and Shuen (1997) have pointed out that it is difficult to calibrate individual contributions to a joint effort and have commented that high-powered incentives might well be destructive of cooperative activity and learning.

DISCUSSION

In her assessment and review of agency theory, Eisenhardt (1989) sets out the main elements of positive agency theory in a table. We repeat this below in Table 2, adding a third column which identifies the areas where behavioral agency theory departs from the standard principal-agent model. According to Eisenhardt the key idea of agency theory is that principal agent relationships should reflect the efficient organization of the costs of information and risk-bearing. The unit of analysis is the contract between principal and agent. The main assumptions are that executives are rational (but see footnote 3), self-interested and risk averse, that there is partial goal conflict between stakeholders, that information is incomplete and not equally shared, and that the overriding organizational objective is efficiency. The problems addressed by the theory involve moral hazard, adverse selection and

how best to share risk, especially where principals and agents have partially differing goals and risk preferences. Proposed solutions to the problems include monitoring through effective corporate governance and outcome-based incentive contracts.

Behavioral agency theory departs from the positive agency framework in three main respects. First, while positive agency theory focuses on the implications for the firm of costs which arise out of the principal-agent relationship, using efficiency as the main assessment criterion, behavioral agency theory focuses on the relationship between agency costs and performance, using efficiency and effectiveness as the yardsticks. The objective of an agency contract is to optimize job performance given the constraints of agency costs. This is achieved at the inflection point λ on the agent's proforma pay-effort curve. Secondly, while agency theory assumes that agents are rational, risk averse and rent seeking, and that there is no non-pecuniary agent motivation, behavioral agency theory proposes a more sophisticated model of man whereby agents are boundedly-rational, loss, risk and uncertainty averse, and where there is a trade-off between intrinsic and extrinsic rewards. Thirdly, while agency theory assumes a linear relationship between pay and motivation, behavioral agency theory proposes a more complex pay-effort function which is affected by loss, risk and uncertainty aversion, the hyperbolic discounting of deferred rewards, inequity aversion, and the trade-off between intrinsic and extrinsic motivation.

INSERT TABLE 2 ABOUT HERE

Implications of Behavioral Agency Theory for the Design of Incentives

Much of the current design thinking about executive compensation ignores behavioral issues and does not take account of agents' preferences, instead falling into the trap of

institutional isomorphism (Di Maggio & Powell, 1983), either in the name of "best practice" by following what other firms do (mimetic isomorphism), or by uncritically doing as regulators say (coercive isomorphism). Behavioral agency theory goes against the current fashion, pointing instead to simpler, more balanced reward systems and more straightforward performance measures. In particularly, contrary to the logic of agency theory, we argue that high powered incentives are not an efficient and effective way of motivating agents. It is not possible to construct an incentive contract for an agent or set performance measures which incorporate all the principal's current objectives and are flexible enough to deal with all possible exogenous shocks which might occur during the performance cycle.   Knowledge constraints resulting from an agent's bounded rationality mean that designing very complex incentive contracts in order to tie the principal's and agent's interests as tightly as possible is likely to have an adverse effect of the agent's job satisfaction and work motivation. Furthermore, at high levels of compensation, crowding-out means that intrinsic motivation which is forgone because of an increase in incentives can only be compensated for by proportionately greater increases in extrinsic rewards. Finally, deferred pay, frequently advocated as a solution to the problem that high levels of executive compensation are seen to be undesirable as a matter of public policy, is in practice an expensive way of paying agents when seen in the context of agent motivation. These arguments are consistent with the "strength of weak incentives" thesis, described above, as advocated by Roberts (2010). They contradict the normative arguments of financial economists who advocate the use of high-powered incentives as a partial remedy for the agency problem (Jensen & Murphy, 1990); see also Bebchuk and Fried (2004: 72).

We argue that, for any group of agents comprising a top management team, there is a balanced set of first best reward strategies, being combinations of fixed and variable pay, contingent and discretionary bonuses, and short and long-term incentives, which allow the

relationship between reward costs, agent motivation and job performance to be optimized. In order to maximize firm performance the selected strategy must be compatible with the strategies selected for other agents in the principal's top management team. Identifying these reward strategies is not a simple matter, ideally requiring an understanding of individual differences between agents in terms of their tolerance of risk and inequity and in the way that they discount future rewards. Partly as a result of the complexity involved in designing appropriately simple incentive and reward systems, ex post discretionary payments to agents may sometimes be warranted as partial gift exchanges in the expectation that they will result in reciprocal gifts of effort (Akerlof, 1982).

Contribution

Agency theory is a central component of the modern theory of the firm (Jensen, 2000; Roberts, 2004). We have explained that the standard theory of agency has significant shortcomings, especially in its failure in practice to explain the relationship between executive compensation, agent behavior and firm performance. While there is, after Cyert and March (1963 |1992), an extensive literature on the behavioral theory of the firm (see Gavetti, Greve, Levinthal & Ocasio, 2012), we do not yet have a satisfactory behavioral agency theory. This paper takes a significant step in correcting this omission. In particular, it advances a theory of behavioral agency which better explains the mechanisms which connect incentives, agent behavior, and the type of high performance outcomes which shareholders desire. This is an important framework, especially for scholars studying executive compensation.

Positive agency theory, like many theories which have their origins in neoclassical economics, aims to provide accurate predictions about economic phenomena without claiming that its foundational assumptions realistically describe the underlying behavioral

processes (see Friedman, 1953). Wakker (2012), who in turn cites Harré (1970), calls this paramorphism. This approach is self-evidently flawed when neither the predictions nor the underlying processes match reality. Behavioral agency theory, on the other hand, aims to explain economic phenomena by reference to descriptions of underlying processes which do match reality. Wakker (2012) calls this homeomorphism. Behavioral economists argue that homeomorphism is more likely to generate useful explanations of actual economic phenomena and hence is a better approach to theory building.

Part of the validity which we claim for behavioral agency theory is based on the way in which it adapts and integrates existing theory. Steel and König have emphasized the importance of consilience in theory development, arguing that: "if a theory can be shown to have consilience, its scientific validity is vastly improved, since it represents different avenues of inquiry coming to similar conclusions" (Steel & Konig, 2006: 889). A major contribution of this paper lies in the way in which it integrates a number of different literatures: in particular, the neoclassical economic theory of agency (after Jensen & Meckling, 1976); work motivation theory (for example Locke & Latham, 1984; Steel & Konig, 2006; and Vroom, 1964); the literature on choices, values, heuristics and biases (after Kahneman & Tversky, 1979; and Tversky & Kahneman, 1992); and the upper echelons approach to strategic leadership (after Hambrick & Mason, 1984). Our paper also makes a contribution to the management literature by updating Eisenhardt's (1989) review of agency theory for management scholars, incorporating new ideas from behavioral economics. In addition, and significantly, the paper has important implications for practice in the way that it advocates the use of balanced executive reward strategies and weak incentives.

Conclusion

Formally, behavioral agency theory comprises four inter-connected equations, two figures, ten propositions, and a supporting narrative. Equation (1), after Boxall and Purcell (2003), connects an agent's job performance with his or her ability, motivation and opportunity set. Equation (2), after Vroom (1964), which is in turn modified by the inclusion of time discounting, risk and loss aversion to become Equation (3), after Steel and König (2006), explains the relationship between compensation and agent motivation. Equation (4) distinguishes between intrinsic and extrinsic motivation and identifies a potential trade-off between the two. Figure 1 explains the place of agent performance and work motivation in the firm's performance cycle and Figure 2 illustrates an agent's pay-effort curve.

In this paper we have sought to provide a better understanding of the micro-foundations of agency theory, especially as it applies to executive compensation, based on a more realistic set of assumptions about agent behavior. We hope that others will join us in further developing behavioral agency theory, in testing it empirically, and in identifying other implications for business practice.

REFERENCES

Adams, J. S. 1965. Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology*: Academic Press, New York.

Ainslie, G. 1991. Derivation of 'Rational' Economic Behavior from Hyperbolic Curves. *American Economic Review, 81(2)*: 334-340.

Ainslie, G., & Haslam, N. 1992. Hyperbolic discounting. In G. Loewenstein and N. Haslam (Eds.), *Choices over time*. New York: Russell Sage Foundation.

Akerlof, G. 1982. Labor contracts as partial gift exchange. *Quarterly Journal of Economics, 97*: 543-569.

Alchian, A., & Demsetz, H. 1972. Production, information costs and economic organization. *American Economic Review, 62*: 777-795.

Arrow, K. 1985. The economics of agency. In J. Pratt and R. Zeckhauser (Eds.), *Principals and agents: the structure of business*. Cambridge, MA: Harvard Business School Press.

Balkin, D. 2008. Explaining High US CEO Pay in a Global Context: An Institutional Approach. In L. Gomez-Mejia and S. Werner (Eds.), *Global Compensation: Foundations and Perspectives*. London: Routledge.

Barnard, C. 1968. *The functions of the executive*. Cambridge, MA: Harvard University Press.

Bebchuk, L., & Fried, J. 2004. *Pay without performance – the unfilled promise of executive compensation*. Cambridge, Massachusetts: Harvard University Press.

Bebchuk, L., Fried, J., & Walker, D. 2002. Managerial power and rent extraction in the design of executive compensation. *University of Chicago Law Review, 69*: 751 – 846.

Becker, G. 1993. *Human capital. A theoretical and empirical analysis with special reference to education* (Third ed.). Chicago: University of Chicago Press.

Besley, T., & Ghatak, M. 2005. Competition and incentives with motivated agents. *American Economic Review, 95*: 616-636.

Boivie, S., Lange, D., McDonald, M., & Westphal, J. 2011. Me or we: the effects of CEO organizational identification on agency costs. *Academy of Management Journal, 54(3)*: 551-576.

Boxall, P., & Purcell, J. 2003. *Strategy and human resource management*. Basingstoke, Hants: Palgrave Macmillan.

Bratton, W. 2005. The Academic tournament over executive compensation. *California Law Review, 93(5)*.

Camerer, C., Loewenstein, G., & Rabin, M. 2004. *Advances in Behavioral Economics*. Princeton, NJ: Princeton University Press.

Carpenter, M., Geletkanycz, M., & Sanders, W. 2004. Upper echelons research revisited: antecedents, elements, and consequences of top management team composition. *Journal of Management, 30(6)*.

Charreaux, G. 2002. Positive agency theory: place and contributions. In E. Brousseau and J. Glachant (Eds.), *The economics of contracts*. Cambridge, England: Cambridge University Press.

Combs, J., & Skills, M. 2003. Managerialist and Human Capital Explanations for Key Executive Pay Premiums. *Academy of Management Journal, 46(1)*: 63-73.

Cyert, R., & March, J. 1963 |1992. A behavioral theory of the firm. Cambridge, MA: Blackwell.

Deci, E., & Ryan, R. 1985. *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Devers, C., Cannella, A., Reilly, G., & Yoder, M. 2007. Executive compensation: a multidisciplinary review of recent developments. *Journal of Management, 33*: 1016-1072.

Di Maggio, P., & Powell, W. 1983. The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*: 147-160.

Eisenhardt, K. M. 1989. Agency theory: an assessment and review. *Academy of Management Review, 14*: 57-74.

Fehr, E., & Schmidt, K. 1999. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*: 817-868.

Finkelstein, S., Hambrick, D., & Cannella, A. 2009. *Strategic leadership: theory and research on executives, top management teams, and boards*. Oxford: Oxford University Press.

Foss, N. 2010. Bounded rationality and organizational economics. In P. Klein and M. Sykuta (Eds.), *The Elgar companion to transaction cost economics*. Cheltenham: Edward Elgar Publishing Limited.

Frederick, S., Loewenstein, G., & O'Donoghue, T. 2002. Time Discounting and Time Preference: A Critical Review. *Journal of Economic Literature, 40(2)*: 351-401.

Frey, B. S., & Jegen, R. 2001. Motivation crowding theory. *Journal of Economic Surveys, 15*: 589-611.

Friedman, M. 1953. The Methodology of Positive Economics. In M. Freidman (Ed.), *Essays in Positive Economics*. Chicago: University of Chicago Press.

Frydman, C., & Jenter, D. 2010. CEO compensation. *Annual Review of Financial Economics, 2*: 75-102.

Gabaix, X., & Landier, A. 2008. Why has executive pay increased so much? *Quarterly Journal of Economics, 123*: 49-100.

Gavetti, G., Greve, H., Levinthal, D., & Ocasio, W. 2012. The behavioral theory of the firm: assessment and prospects. *The academy of management annals*. New York: Routledge Taylor & Francis Group.

Gomez-Mejia, L., Berrone, P., & Franco-Santos, M. 2010. *Compensation and organizational performance - theory, research and practice*. Armonk, NY: M.E. Sharpe, Inc.

Gomez-Mejia, L., & Wiseman, R. 1997. Reframing executive compensation: an assessment and outlook. *Journal of Management, 23*: 291-374.

Graves, S., & Ringuest, J. 2012. Patient Decision Making: Exponential versus Hyperbolic Discounting. *Managerial and Decision Economics*: Wiley Online Library.

Grossman, S., & Hart, O. 1983. An Analysis of the Principal-Agent Problem. *Econometrica, 51(1)*: 7-45.

Hambrick, D. 1994. Top management groups: a conceptual integration and reconsideration of the "team" label. In B. Staw and L. Cummings (Eds.), *Research in Organizational Behavior*: 171-213. Beverly Hill: JAI Press.

Hambrick, D., & Mason, P. 1984. Upper echelons: the organization as a reflection of its top managers. *Academy of Management Review, 9(2)*: 193-206.

Harre, R. 1970. *The Principles of Scientific Thinking*. London: Macmillan.

Hart, O. 1995. *Firms, Contracts and Financial Structure*. Oxford: Oxford University Press.

Jensen, M. 1983. Organization theory and methodology. *The Accounting Review, 58*: 319-339.

Jensen, M. 1998. *Foundations of organizational strategy*. Cambridge, MA: Harvard University Press.

Jensen, M. 2000. *A theory of the firm: governance, residual claims, and organizational forms*. Cambridge, MA: Harvard University Press.

Jensen, M., & Meckling, W. 1976. Theory of the firm: managerial behaviour, agency costs and ownership structure. *Journal of Financial Economics, 3*: 305-360.

Jensen, M., & Meckling, W. 1994. The nature of man. *Journal of Applied Corporate Finance, 7(2)*.

Jensen, M., & Murphy, K. 1990. Performance pay and top-management incentives. *Journal of Political Economy, 98*: 225-264.

Kahneman, D., & Tversky, A. 1979. Prospect theory – an analysis of decision under risk. *Econometrica, 47*: 263-291.

Kreps, D. 1997. Intrinsic motivation and extrinsic incentives. *The American Economic Review, 87*: 359-364.

Lazear, E. P., & Rosen, S. 1981. Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy, 89*: 841 – 864.

Leibenstein, H. 1966. Allocative efficiency vs. x-efficiency. *The American Economic Review, 56*: 392-415.

Lepper, M., & Greene, D. 1978. Introduction. In M. Lepper and D. Greene (Eds.), *The hidden costs of reward - new perspectives on the psychology of human motivation.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Lewin, K. 1938. The conceptual representation and measurement of psychological forces. *Contributions to Psychological Theory, 4*: 247.

Locke, E., & Latham, G. 1984. *Goal setting - a motivational technique that works*. New Jersey: Prentice-Hall Inc.

Locke, E., & Latham, G. 1990. *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.

Locke, E., & Latham, G. 2002. Building a practically useful theory of goal setting and task motivation - a 35 year odyssey. *American Psychologist, 57*: 705-717.

Michelman, F. R. 1967. Property, utility, and fairness: comments on the ethical foundation of 'just compensation' law. *Harvard Law Review, 80*: 1165-1257.

Mintzberg, H. 1997. *The nature of managerial work*. London: Prentice Hall.

Mintzberg, H. 2009. *Managing*. London: FT Prentice Hall.

Pepper, A. 2006. *Senior executive reward – key models and practices*. Aldershot: Gower.

Pratt, J., & Zeckhauser, R. 1985. Principals and agents: an overview. In J. Pratt and R. Zeckhauser (Eds.), *Principals and agents: the structure of business*. Cambridge, MA: Harvard Business School Press.

Rebitzer, J., & Taylor, L. 2011. Extrinsic rewards and intrinsic motives: standard and behavioral approaches to agency and labor markets. In O. Ashenfelter and D. Card (Eds.), *Handbook of labor economics*: 701-772. Amsterdam: North-Holland.

Roberts, J. 2004. *The modern firm*. Oxford: Oxford University Press.

Roberts, J. 2010. Designing incentives in organizations. *Journal of Institutional Economics, 6*: 125-132.

Roberts, J. 2011. Weak incentives: when and why. In D. Marsden (Ed.), *Employment in the lean years*: 185-198. Oxford: Oxford University Press.

Ross Ashby, W. 1956 |1976. *Introduction to cybernetics*. London: Methuen.

Ross, S. 1973. The economic theory of agency: the principal's problem. *American Economic Review, 63*: 134-139.

Sanders, G., & Carpenter, M. 2003. A behavioral agency theory perspective on stock repurchase program announcements. *Academy of Management Journal, 46(3)*: 160-178.

Savage, L. 1954. *The foundations of statistics*. New York: Wiley.

Shafir, E., Diamond, P., & Tversky, A. 1997. Money illusion. *The Quarterly Journal of Economics, 112*: 341-371.

Simon, H. 1945 |1997. *Administrative behavior* (Fourth ed.). New York: The Free Press.

Simon, H. 1957 |1982. The compensation of executives; first published in Sociometry (1957) Vol. 20. *Models of bounded rationality, volume 2: behavioral economics and business organisation* 47-48. Cambridge, MA: The MIT Press.

Sliwka, D. 2007. Trust as a signal of a social norm and the hidden costs of incentive schemes. *American Economic Review, 97(3)*: 999-1012.

Spence, M. 1973. Job market signalling. *Quarterly Journal of Economics, 87(355-374)*.

Spence, M., & Zeckhauser, R. 1971. Insurance, information and individual action. *American Economic Review, 61*: 380-387.

Steel, P., & Konig, C. 2006. Integrating theories of motivation. *Academy of Management Review, 31*: 889 – 913.

Teece, D., Pisano, G., & Shuen, A. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal, 18(7)*.

Tosi, H., & Gomez-Mejia, L. 1989. The decoupling of CEO pay and performance: an agency theory perspective. *Administrative Science Quarterly, 34*: 169-189.

Tosi, H., Werner, S., Katz, J., & Gomez-Mejia, L. 2000. How much does performance matter? A meta-analysis of CEO pay studies. *Journal of Management, 26*: 301-339.

Tversky, A., & Kahneman, D. 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty, 5*: 297-323.

Ungson, G., & Steers, R. 1984. Motivation and Politics in Executive Compensation. *Academy of Management Review, 9(2)*: 313-323.

Vroom, V. H. 1964. *Work and motivation*. New York: Wiley.

Wade, J., O'Reilly, C., & Pollock, T. 2006. Overpaid CEOs and Underpaid Managers: Fairness and Executive Compensation. *Organizational Science, 17(5)*: 527-544.

Wakker, P. 2012. *Prospect theory*. Cambridge: Cambridge University Press.

Wiseman, R., & Gomez-Mejia, L. 1998. A behavioral agency model of managerial risk taking. *Academy of Management Review, 23*: 133-153.

FOOTNOTES

[1] Some commentators (e.g., Roberts, 2011) imply that Jensen and Murphy's empirical evidence is not contrary to agency theory, but suggests instead that it means the (normative) recommendations of agency theory have not been followed in practice. An argument in this form, implying that the absence of two factors (incentive pay and high performance) can be interpreted as evidence of a causal connection between the two phenomena (so that more of the first factor will necessarily lead to more of the second) is hardly justified. It also appears to confuse the positive theory of agency (which should be capable of explaining the world as it is) with normative theory. In practice, as argued long ago by Herbert Simon in 1957 and demonstrated empirically by Gabaix and Landier in 2008, CEO pay is much more closely correlated with company size than company performance.

[2] Jensen (1983) identifies two different strands in the literature on agency theory. He calls these the "positive theory of agency" and the "principal-agent" literature. Eisenhardt (1989) describes the latter as a "general theory of the principal-agent relationship", while Wiseman and Gomez-Mejia (1998) call it "normative agency theory". Positive agency theory focuses on the special case of the principal-agent relationship between owners and managers of large corporations (Charreaux, 2002; Eisenhardt, 1989; Jensen, 1983). Normative agency theory aims to provide a formal theory of the principal-agent relationship in all its guises, including employer-employee, lawyer-client, buyer-supplier etc., (Eisenhardt, 1989).

[3] Eisenhardt (1989) states that positive agency theory also assumes bounded rationality, but we can find no other reference to this in the agency theory literature. After the first, formative papers on agency theory, Jensen and Meckling (1994) later develop the resourceful, evaluative, maximizing model of man (REMM) which they say is consistent with agency theory, but this is still a rational choice model. They subsequently develop a second framework, the pain avoidance model (PAM), but they do not seek to integrate this into agency theory.

[4] Christen et al., (2006) point out that motivation (wanting to work hard) is not the same as effort (working hard and, in doing so, expending time and energy). However, in much the same way that revealed preference is a marker of mental preference, so effort can be thought of as a marker of motivated behavior (see Martin and Tesser, 2009). This means that, in the absence of coercion, effort can be taken to imply the presence of motivation even if motivation does not necessarily result in the expenditure of effort.

[5] Jensen and Meckling do not explicitly mention expenditure on incentives and rewards, i.e., the actual cash costs of incentivizing and rewarding agents, although such expenditure would seem self-evidently to be part of

the cost of agency. Incentive and reward costs can be further broken down into the costs of providing incentives and rewards in the optimal form and mix plus any additional costs incurred in incentivizing and rewarding agents in a way which is sub-optimal. In order to be precise we use the following terminology in this article: total compensation or pay (ω) is the sum of fixed pay and variable pay; variable pay is itself the sum of incentives (awarded ex ante) and discretionary rewards (awarded ex post). "Compensation" and "pay" are treated as synonyms. Use of the sign "ω" follows the convention in labor economics of taking "ω" as the symbol for wages.

[6] There are many different definitions of "bounded rationality". We follow Williamson, who explains that rationality is subject to neuro-physiological rate and storage limits on the powers of agents to receive, store, retrieve, and process information without error (after Williamson, 1975, p.21). Williamson also talks about a further element of bounded rationality, which he calls "language limits", being the constraints on individuals to communicate comprehensively in such a way that they are fully understood by others, but this element is not really relevant to the current article. Foss (2010) provides an elegant summary description of bounded rationality, which he describes in terms of (1) limitations in the human capacity to process information; (2) attempts to economize on mental effort by relying on short-cuts or heuristics; and (3) a consequence of the fact that cognition and judgement are subject to a wide range of biases and errors.

[7] In this paper we largely ignore the Knightian distinction between risk (probabilistic outcomes) and uncertainty (indeterminate outcomes), instead treating "risk" and "uncertainty" as synonymous.

[8] Note however that Barnard used the term "efficiency" in an entirely different sense: to Barnard an organization is "efficient" if it satisfies the motives of its members.

[9] Expectancy, a measure of probability, takes values between 0 and +1. Instrumentality takes values between +1 (meaning it is believed that the first outcome will certainly lead to the second outcome) and -1 (meaning it is believed that the second outcome is impossible in the event of the first outcome).

[10] Temporal motivation theory combines expectancy and instrumentality into one operator, which Steel and König call "expectancy" but which is essentially the same thing as subjective probability after Savage (1954). While this loses some of the richness of Vroom's conceptualization of expectancy and instrumentality (especially the possibility that instrumentality may be negative) it is a pragmatic simplification of the theory and hence is followed here.

[11] A more complex way of representing Steel and König's motivation function, which distinguishes between gains and losses and hence accounts for loss aversion is:

$$X_i = \sum \left\{ \frac{E_{ik}^+ \times V_k^+}{1 + \delta^+ t} \right\} - \sum \left\{ \frac{E_{ik}^- \times V_k^-}{1 + \delta^- t} \right\}$$

This expression of the formula explicitly recognizes that the expectancy, valence and the average discount factor will differ for gains (represented by $^+$) and losses (represented by $^-$).

TABLE 1: Assumptions about the Nature of Man under Positive Agency Theory and

Behavioral Agency Theory

Assumption	Economic man	Behavioral economic man
Principal's risk preference	Principals are risk neutral	As for agency theory
Agent's utility function	Agents are rent seeking, agent's utility is positively contingent on pecuniary incentives and negatively contingent on effort	As for agency theory, but subject to constraints relating to rationality, motivation, loss, risk, uncertainty and time preferences
Agent's rationality	Agents are rational	Agents are boundedly rational, i.e., subject to neuro-physiological rate and storage limits on the powers of agents to receive, store, retrieve, and process information without error
Agent's motivation	There is no non-pecuniary agent motivation	Motivation is both intrinsic and extrinsic. Intrinsic and extrinsic motivation are neither independent nor additive.
Agent's risk preference	Agents are risk averse	Agents are loss averse below a gain/loss inflection point; otherwise risk averse
Agent's time preferences	Agents' time preferences are calculated according to an exponential discount factor	Agents' time preferences are calculated according to a hyperbolic discount factor
Agent's preference for perceived equitable pay	Not defined	Agents are inequity averse

TABLE 2: Overview of Positive Agency Theory vs. Behavioral Agency Theory

	Positive agency theory (after Eisenhardt, 1989)	**Behavioral agency theory**
Key idea	The primary importance of aligning the interests of principals and agents. The principal-agent relationship should reflect efficient management of the costs of information and risk-bearing	The primary importance of agent performance and work motivation. The principal-agent relationship should reflect the efficient *and effective* management of the relationship between executive compensation, firm performance and shareholder interests
Unit of analysis	Contract between principal and agent	Contract between principal and agent
Human assumptions	Agents are rational, self-interested, risk averse	Agents are boundedly rational, loss, risk and uncertainty averse, hyperbolic time discounters, inequity averse, and there is a trade-off between intrinsic and extrinsic motivation
Organizational assumption	Partial goal conflict between principals and agents, efficiency as the main performance criterion, information asymmetry	Partial goal conflict between principals and agents, efficiency *and effectiveness* as the main performance criterion, information asymmetry
Information assumption	Asymmetric information and incomplete contracting	As for agency theory; goal setting used as a pragmatic solution to information asymmetry
Primary factor(s) determining the principal-agent relationship	The principal's wish to align the agent's objectives with the principal's owns objectives (alignment)	The principal's wish to align the agent's objectives with the principal's own objectives (alignment) and to motivate agents to give high performance, given their abilities and opportunities (motivation)
Contracting problems	Moral hazard and adverse selection	As for agency theory

Key mechanisms	Monitoring and incentive contracts	As for agency theory, except that incentive contracts can also help to meet the motivation objective
Problem domain	Where principals and agents have different goals and risk preferences e.g., regulation, compensation, vertical integration, transfer pricing	As for agency theory; especially relevant to executives and executive compensation

FIGURE 1
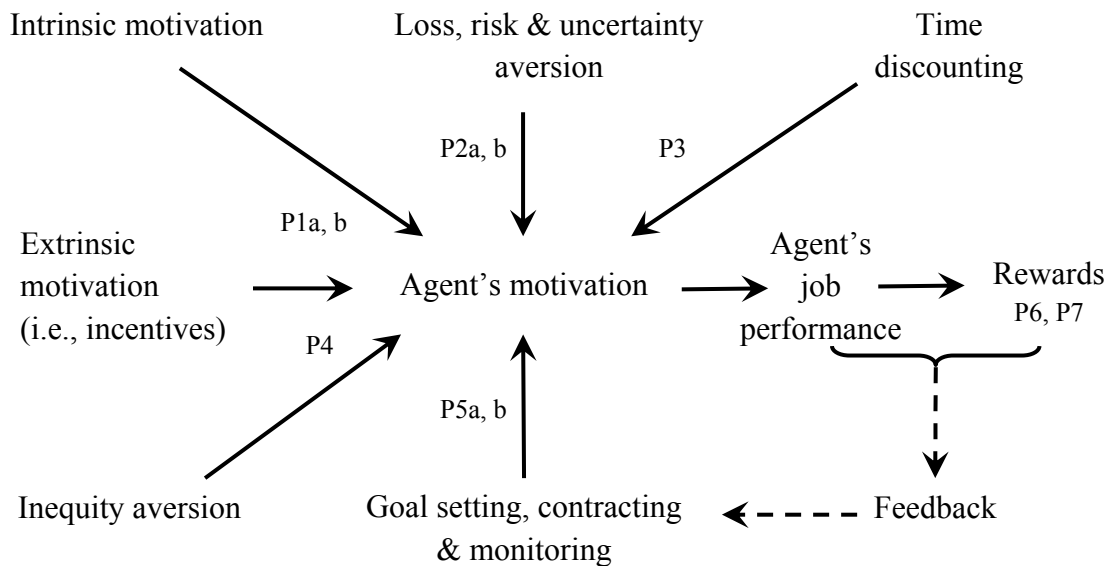**Agent's Job Performance and Work Motivation Cycle**

FIGURE 2
Agent's Pay-Effort curve