The dark side of trust: when trust becomes a ‘poisoned chalice’.

Abstract

Trust is typically portrayed as inherently beneficial to all concerned; something which is inherently good. In this paper we focus on interpersonal trust and argue that, while there are undoubted benefits, trust can also be problematic; there are circumstances in which it can become a ‘poisoned chalice’ for one or other of the parties involved. We question whether the potential for negative experiences in relation to trust has been fully explored and argue that its true dark side lies within the dynamics of the trust relationship. From this stance we use the Social Exchange and gift giving literatures to re-appraise trust in a way that highlights the importance of expectation and intent and demonstrate the circumstances in which trust may not be beneficial for one of the parties involved. We conclude with a research agenda which we believe will further develop our understanding of this complex human interaction.

Keywords (alphabetical order)

‘Dark side of trust’, manipulation, trust relationship, trust and gift-giving, unwanted obligation.

Introduction.

Trust has long been depicted as an inherently positive experience and a desirable quality which should be sought after in relationships (Fukuyama, 1995). Interpersonal trust is considered a vital social resource that can facilitate cooperation and enable co-ordinated social interaction and co-operation (Brenkert, 1998a; Dirks and Ferrin, 2001). In short, highly normative assumptions hailing trust’s myriad benefits dominate the discourse: what Gargiulo and Ertug call an ‘optimistic bias’ (2006: 165). For, as Sievers observes (2003: 356), ‘how could anyone be against
trust per se? It is – like motherhood and apple pie – a good thing and a necessary constituent of the social fabric’.

We are concerned that this litany has become taken-for-granted, almost to the point of being unchallengeable, and that trust may be seen as a catch-all panacea (McAllister, 1997). In 1998 Flores and Solomon observed the danger that trust was being oversold and we believe that danger still exists. While there are undoubted benefits connected with trust, we argue that its virtues are only part of the picture and, as with most organisational phenomena, trust has its downside. We therefore explore in this paper what might distinguish a negative experience of interpersonal trust from a positive experience. In order to achieve this, we examine what is commonly portrayed as the ‘dark side’ of trust, identify the limits to current thinking and offer an alternative view.

Our primary contribution in this paper is to show how the particular nature of trust and the behavioural strictures implied in trust dynamics – the same ones which generate the positive experiences lauded in the literature – can also facilitate negative dynamics and outcomes. We argue that trust’s dark side does not materialise when it turns out to be misplaced, which is the most common depiction of negative trust experiences instead we suggest that it arises when the dynamics of interpersonal trust draw the focal party (whether giver or recipient) into an uncomfortable exchange dilemma from which it is difficult to extricate oneself; where the very nature of trust means that most of the options available as a response are neither viable nor attractive.

Some of these uncomfortable scenarios can be unintentional consequences of genuine, sincere trust dynamics, but there is also the potential for those dynamics to be manipulated deliberately. In either case, trust can become the ‘poisoned chalice’ of our title; something that has the appearance of being good (beneficial, virtuous) on the surface but which, in reality, can lead to negative experiences and unwanted outcomes for the focal party. The implications of our analysis challenge
several prevailing assumptions governing this ‘central, superficially obvious but essentially complex’ concept (Blois, 1999: 197).

The paper proceeds as follows: we first define trust and explain its essential dynamics. We then note the scant treatment that its dark side has received to date, before turning to social exchange theory (SET) and the anthropological gift-giving literature to provide the primary theoretical basis for our own analysis of this issue. We highlight several insights from these two literatures which facilitate a broader understanding of what the ‘dark side’ of trust means. Throughout we use vignettes from organisational life to illustrate when trust might become a poisoned chalice. In the discussion and conclusion, we consider the implications of our analysis for trust relationships, including for several ongoing debates in the literature regarding the nature of trust.

Trust

Before we can analyse what it is about the process of trusting that can render trust either a benign experience or a poisoned chalice we first need to understand the process of trusting and being trusted. Commonly the process is described as follows; parties to any trusting interaction (‘A’ and ‘B’) exchange small-scale concessions or offerings in the form of trust’s risk-taking act. Party A’s trust bestowed upon B is either vindicated (B elects not to exploit A’s vulnerability, fulfils her/his obligations competently and honestly, and possibly reciprocates in kind), in which event trust may develop further. Alternatively, B chooses to take advantage of A’s vulnerability or is incompetent in discharging her/his implied obligations, undermining trust to the point where it may disappear from the relationship entirely.

In any trust relationship, there is a giving party (the trustor) and a receiving party (the trustee). Trust becomes salient in a situation entailing a degree of uncertainty and risk for one or both parties, whether the trustor or the trustee. In fact the relevance of trust is due to the uncertainties regarding the actions of others in a
situation of dependency (see Luhmann 1979) and as Möllering (2007: 8) notes this uncertainty is irreducible as ‘the agency of both trustor and trustee, who are autonomous in that their states of mind and actions are not fully determined, the result being that neither manifestations of trust nor the honouring of trust can ultimately be forced or guaranteed’ As we shall see, the inherent ambiguity in trusting situations is an important factor in the context of the dark side.

Trust is therefore a process, it is not simply a psychological state, or a cognitive, calculative orientation toward risk (Kramer, 1999, Rousseau et al., 1998). This is vital for our argument. The trust process consists of three stages (Dietz and den Hartog, 2006; McEvily et al, 2003) beginning with a set of beliefs about the other party’s trustworthiness, commonly understood to comprise assessments of their ability, benevolence and integrity (Mayer et al., 1995). A decision to trust is taken on the basis of those beliefs, summarised in Rousseau et al’s widely adopted definition of trust as ‘a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behaviour of another’ (1998: 395); what Giddens (1990) has described as a leap of faith. Distrust occurs when the trustor has confident negative expectations about the other party and, based on these, decides not to accept vulnerability (Lewicki et al.,1998). The final stage, arising from the decision, is a risk-taking act, undertaken to demonstrate one’s trust (Mayer et al., 1995). Such acts include increased collaboration and reliance, sharing scarce or valuable resources and sensitive information, and deliberately reduced monitoring (Dietz and den Hartog, 2006). Note that Rousseau et al’s definition omits the behavioural manifestation of trust, while in Mayer et al.’s model, it is viewed as a necessary outcome of trust, but not part of trust itself. We disagree. As we will show, the risk-taking act is trust’s defining stage, and the way that it is both sought and enacted (or not) can lead to the manifestation of trust’s dark side.

If viewed as a process, it is apparent that trust is subject to the outcomes of social interactions between the parties concerned (Lewicki et al., 2006; Zand, 1972).
Its presence (and maintenance), or its absence (or threat thereof), emerges from, and defines, the relationship between parties. Fundamentally, trust is relational (Schoorman et al., 2007). Trust can be offered, requested, bestowed or denied, and the moment of the exchange is the risk-taking act. The content and progress of exchanges determine whether trust is positively or negatively experienced, even superseding the cognitions of the parties involved. Flores and Solomon (1998: 206) argue that trust is a ‘social practice, defined by choices’, it evolves in relationships and is created (destroyed) through everyday exchanges in which promises, commitments and the frustration of such promises and commitments occur. The process of trusting involves powerful emotions and values (Rousseau et al., 1998). These emotions are typically portrayed as positive (Frederickson and Joiner, 2002) such as reassurance and pride while not trusting or being trusted is associated with negative emotions such as anxiety or humiliation. People do tend to prefer trust in their relationships over distrust (Graebner, 2009) but we question whether this optimistic portrayal of trust’s affective state always holds true.
Current views on the ‘dark side’ of trust

In the context of human activity and interaction there are few things which are wholly positive yet trust’s ‘dark side’ has received surprisingly little attention within the dedicated trust literature. In neither Mayer et al’s seminal conceptual paper (1995), nor their anniversary reflection (Schoorman et al., 2007), is any downside to trust considered beyond distrust and trust violations. Kramer, in his widely cited review (1999), noted several barriers to trust but offered few reservations about the experience of trust other than distrusting scenarios. He has since argued, in the wake of the Bernie Madoff ponzi scandal, that people may be trusting too much and advocates practical steps for managers to be vigilant against biases that may render them a victim of manipulated trust (Kramer, 2009).

Where it is addressed, trust’s dark side is most commonly linked with poor judgement. For example, Gargiulo and Ertug (2006) identify what they call the ‘dark side’ of trust as occurring when the trustor strays beyond a critical threshold of confidence such that her trust in another becomes inappropriate and ill-judged. Similarly, Currall and Epstein’s (2003) trust-derived interpretation of the Enron scandal notes that too many people showed excessive trust in the firm’s senior management. McEvily et al. (2003: 99-100) allude to trust’s ‘considerable downside’, noting how it can ‘lead the trustor astray and ‘may produce systemic biases that can result in judgements that are substantially flawed and costly’. This downside has been demonstrated empirically in Langfred’s (2004) study of complacent self-managing teams, and Patzelt and Shepherd’s (2008) study of underperforming alliances.

Yet, what these authors have highlighted is the possibility of mistaken judgement: situations in which trust may be misplaced, and hence counter-productive for the trust giver. This misunderstands a fundamental aspect of trust, which is that the possibility of betrayal or disappointment is inherent to the experience and hence unavoidable in any trusting encounter. Indeed, it is one of trust’s defining conditions
as without this possibility there would be little risk. The prospect of getting a trust decision wrong is certainly trust’s regrettable side, but we question whether it is truly a dark side, a phrase which implies something invidious or even sinister.

We are aware of a limited number of overtly sceptical takes on trust that draw attention to the possibility of trust being used in a manner which does not match with the idea that it is an inherently beneficial experience. Gambetta (1988) was among the first to point to bonds of trust being problematic in certain circumstances, citing the powerful bond among mafiosi. As well as considering the risks from misattributed trust, McAllister (1997: 93) examined ‘the susceptibility of trust relationships to manipulation and abuse’ and ‘the tendency towards persistence in failed and failing trust relationships’. Flores and Solomon (1998: 208) observe the potential for trust to be utilised as a manipulative tool in business contexts if it is viewed as ‘a lubricant to make an operation more efficient’. Sievers (2003: 360-363) points to the ‘self-deceiving character’ of what passes for trust in today’s organisations: weak, transient, unlikely to be genuine. He suggests that trust, like ‘motivation’ is a managerial ruse for re-establishing order and masking the disorder, misery and despair felt by many at work. He cites Hardy et al.’s (1998) case study of a plant closure, in which the management fooled the trade union with their apparent concern and candour. Importantly for our purposes, Hardy and colleagues argue that the managers offered an outward display of trust to mask a covert, deliberate, use of antagonistic power (for a similar case study, see Timming, 2008).

These sceptical takes begin to highlight what we believe contributes to the dark side of trust. In our view this is not about misplaced trust but occurs when the trust process results in a trusting situation that is unwelcome and/or disadvantageous but effectively unavoidable for one or both parties concerned. As we shall illustrate later, such dark side situations can arise unintentionally or be created deliberately as a result of ‘motivated behaviour by an employee or group of employees that has negative consequences for an individual’ (Griffin and O’Leary-
Kelly, 2004: 4-5). We believe that trust’s dark side is inherent in the very nature of the trust relationship; it is in essence about exchange, reciprocation and obligation.

**Understanding trust as an exchange commodity.**

As we have noted above, trust occurs within the context of a relationship and strengthens or declines as the result of the exchanges which occur but it is clearly different from money or other tangible resources which may be exchanged. Foa and Foa (Coyle-Shapiro and Conway, 2004; Cropanzano and Mitchell, 2005) plotted exchange resources against two dimensions: the concreteness of the resource (its physical tangibility) and particularism-universalism (the importance of the identity of the giver: the extent to which who is giving the resource matters for the resource itself). Trust is clearly intangible; we cannot touch trust, or measure it out in clear quantities. In many incidences, it might not even be observable, though it is demonstrated through a risk-taking act. Trust is more often something that is felt or believed. In Homans’ terms it is a resource of ‘symbolic value’ (1958, cited in Cropanzano and Mitchell, 2005: 890) that is being exchanged. Trusting someone and being trusted are acts charged with meaning – symbolism – for the relationship and herein lies some of the potential for the dark side. Trusting someone implies an endorsement of their character and capabilities, and hence their status, whether in general terms or within a given social network. Thus, trust as a resource is bound up with our identity (Tajfel and Turner, 1985) and as such has the power to create a strong obligation not to disappoint the trust giving party. Indeed, Tyler (2003) links this symbolic value of trust to status and identity information: trust signals to the trust recipient that (s)he is accepted and seen as an important group member. As such trust has the power to create a strong obligation in the recipient not to disappoint the trust giving party.

Trust is also highly particularistic: the identity of the focal party is crucial. If the decision to trust is intimately linked to self-identity, through parties’ chosen
affiliations, it matters who trusts us, and who we choose to trust. The importance of who trusts us may be further heightened in relationships where there is a power imbalance such as between a line manager and employee. Intangible and particularistic resources are, in Foa and Foa's words, 'an expression of evaluative judgement which conveys high or low prestige, regard or esteem' (cited in Coyle-Shapiro and Conway, 2004: 12).

Additionally, a number of authors argue for a profound emotional dimension to the experience of trust: for example, McAllister's (1995) measure of trust explicitly taps this dimension, alongside the more conventional cognitive dimension. It simply feels good to experience the reassurance and comfort from trusting someone (Lewicki et al., 1998), and the pride in being trusted (Pettit, 1995; Salomon & Robinson, 2008). At the same time, not trusting or not being trusted can induce feelings of anxiety, fear, embarrassment, even shame and humiliation (Lewicki et al., 1998). The benefits associated with trusting and being trusted, and the deleterious outcomes associated with distrust, mean that the process is fraught with feelings. Both trust and the refusal to trust and be trusted are thus linked to strong emotional experiences and it is not surprising that the preference tends to be for trust rather than distrust to be part of relationships (Graebner, 2009).

In summary the characteristics of trust are that it is symbolic rather than material, particularistic rather than universal, and socio-emotional rather than instrumental. These are characteristics which have the potential to create and structure obligation within the context of a relationship and which help us to appreciate why trust is such a potent construct (precious and treasured, but also potentially onerous and difficult to resist).

**Trust, reciprocation and obligation: insights from SET and gift-giving.**

Since trust is relational, many authors argue that its nature is intimately bound up in social exchange, and vice versa (see Blau, 1964; Cropanzano and Mitchell,
2005; Whitener et al., 1998). The two are thought to account for the other, and trust is commonly understood to be realised through social exchange processes: ‘reciprocation reinforces and stabilises trust, the axis upon which social exchange resolves’ (Aryee et al., 2002: 271). Social exchange is about the initiation and expansion of exchanges in relationships. For Blau (1964: 91) it is concerned with the ‘voluntary actions [later called ‘investments’] of individuals that are motivated by the returns they are expected to bring, and typically do bring, from others’. Social exchange ‘involves a series of interactions [that] are usually seen as interdependent and contingent on the actions of another person’ (Cropanzano and Mitchell, 2005: 874). Social benefits, like trust, have ‘no exact price’ (Blau, 1964: 96) and hence ‘cannot be bartered lest their value as genuine feelings or judgements be compromised’ (ibid). Moreover, standard governing principles and processes for economic transactions – setting prices; formal contracts; the threat of sanction – may not be available for regulating trust, or may not be an attractive or viable option and herein lies further potential for the dark side.

That said, there are a few suggested guidelines for reciprocity in the SET literature. One is to reciprocate in a manner that is contingent upon the original offer received. The norm is that ‘an individual who supplies rewarding services to another obligates him. In order to discharge this obligation, the second must furnish benefits to the first in turn’ (Blau 1964: 89 – emphasis added). According to Blau, the person who receives a reward from associating with another has an incentive to furnish inducements that continue the association (1964: 100). Gouldner (1960) argued that the return must be roughly equivalent, in order to do justice to the original transaction and to sustain the exchange. Thus, the obligations of social exchange can create a bond that needs regular confirmation in order for a relationship to endure and prosper. Importantly, this means that the norm of reciprocity delimits and even instructs the likely content of the exchanges, and reinforces the enduring social patterns created, by constraining the parties’ options. Indeed, the norm’s powerful
hold on our thinking and behaviour – some argue that the principle of reciprocity is hard-wired into our brains (Cialdini, 2001; Carmichael and MacLeod, 1997) – means that resisting its implied obligations is formidably difficult. For a valued commodity such as trust, this could be expected to intensify the pressure of the obligation (see Coyle-Shapiro and Conway, 2004) to a degree that, we argue, can become coercive.

Trust dynamics are commonly depicted in ways that match those in SET (Whitener et al., 1998) but, we would also argue, there is resonance with those described in the anthropological literature relating to gift-giving. Both the theory of social exchange and that on gift-giving usefully explore issues relating to reciprocity, obligation and patterns of exchange that are similarly understood to exert a powerful influence on people’s thinking and behaviour (Cialdini, 2001). While the norm of reciprocity (Gouldner, 1960; Whitener et al., 1998; Lemmergaard and Muhr, 2011) is important in both literatures, the issue of reciprocity has been more contentious in the context of gift giving where more attention has been paid to intentions which lie behind a gift. A contribution which we believe is particularly helpful when exploring the dark side of trust; in understanding the nature of these exchange relations, the tensions inherent within them, and the potential for negative effects.

Most anthropologists, from Mauss (1954) onwards, have argued that gift-giving carries with it an implied expectation of a return in kind at some point, whether with a tangible object or in a more symbolic fashion (Bergquist and Ljungberg, 2001). However, some have stressed an absence of reciprocity as a defining characteristic expressed in the concept of the ‘free gift’, a gift, which is given without expectations of any return (Derrida 1992; Jones, 2003). Derrida (1992) has argued that a pure gift requires that there should be no reciprocity and no memory of a gift given. However, as Lemmergaard and Muhr (2011) observe, the difficulty that theorists have had in demonstrating the existence of such gifts serves to highlight that, in reality, a gift is never entirely free or reciprocal. Hence many authors in the gift-giving literature would consider that Derrida’s notion of the ‘free gift’ is best understood as an ideal
type in the Weberian sense: an ideal against which social reality is interpreted and measured. We would suggest that the notion of this ‘wanted illusion’ that gift recipients are looking for (Ortmann, 2004) paradoxically may even facilitate gifts being given instrumentally with the intent of creating obligations, even to the extent that gift recipients could be manipulated into situations of dependency. Such an effect is for instance, discussed in the marketing literature where a number of studies show that ‘free gifts’, gifts which are given without explicit requests, are more likely to evoke feelings of obligation than conditional gifts and thus are more likely to lend themselves to manipulation (Bodur and Grohmann, 2005).

Mauss (1954) represented gifts as ambiguous, combining both generosity and self-interest, suggesting that influence is carried in all gifts (Lemmergaard and Muhr, 2011). Gift-givers who are more interested in others’ interests either seek to create mutual benefits or devote resources unselfishly out of a deep concern for the recipient (Uhl-Bien et al., 2003). Under these conditions it is suggested that recipients react with gratitude, which is a positively valenced emotion (Lazarus and Lazarus 1994) and this is linked to pro-social motivation (Gray et al. 2001). Mainly self-interested gift-givers initiate exchanges in order to obtain resources from the recipient, which would be otherwise hard or very costly to obtain. Tsang (2006) suggests that such self-interested intentions create indebtedness on the part of the focal party: that is, an obligation to reciprocate which may be linked to negative emotions (for example Greenberg, 1980) and avoidance motivation (Gray et al., 2001). Self-interested intentions might, therefore, turn trust sour.

In addition to recognising the pernicious effects of self-interested motives, the gift-giving literature analyses manipulating behaviour. Mauss (1954), Schwartz (1967) and Wolfinbarger (1990) all acknowledge self-interest or the engineering of recipient indebtedness as a possible motive behind gift-giving, seeing both tactics as useful for gaining, and maintaining, status and control. Sherry argues that the act of giving is not merely ‘a vehicle of social obligation’ but also for ‘political manoeuvre’
(1983: 157 – emphasis added), thus the gift-giver may instrumentalise gift-giving even to the point where the gift-recipient is exploited and harmed. In other words, they may handle gift-giving opportunistically by misusing their power to their own advantage. Blau (1964: 106-112) also saw the potential for the distribution of social exchange being used as a self-interested power-play.

As we shall discuss shortly, unlike the vast majority of the trust and the SET literature, the gift-giving literature identifies and explores the whole range of possible intentions of a gift-giver from the more altruistic ‘free gift’ to the more sinister manipulative gift. The potential for an unwelcome and/or manipulative side of exchange is recognised as is the reality that, in certain incarnations, gift-giving can carry unwelcome consequences for the giver or the recipient (Lemmergaard and Muhr, 2011).

**Unwelcome obligations**

The gift-giving literature has intensively discussed the anxieties stemming from fears of not meeting the expectations of the other party (Wooten, 2000). If we apply this to the giving and receiving of trust we can see the potential for trust to be problematic for both sides. For example, a gift-giver might worry that their gift does not match the gift-recipient’s preferences (Camerer, 1988) or the pre-existing expectations on the part of the gift recipient might not match the reality of the gift, or the gift might not reflect the recipient’s self-concept correctly (Sunwolf, 2006). In a trust context the trust giver might worry that their trust is misplaced, something which we have argued is inherent to trust and not its dark side. But a scenario of inappropriately placed trust can become a poisoned chalice if requests for that trust are contrary to the trustee’s better judgement and create an uncomfortable obligation. Our first illustrative scenario explains how.
**Scenario 1: Reluctant trust – in which Anthony exhorts Ben to trust him.**

In this scenario, Anthony’s appeal to Ben is that he can, and even should, be trusted. If Ben has ‘confident positive expectations’ that Andrew can be relied upon, then a normal and genuine trust encounter can ensue. That is not a ‘dark side’ scenario. However, if Ben is wary of Anthony, and/or feels that to trust him might be contrary to his own interests, Ben may be thought to have few ‘confident positive expectations’, sensing that placing trust in Anthony may be unwise. The, ‘poisoned chalice’ dilemma for the trust-giver (Ben) occurs because although Anthony’s request for trust ought to be treated with scepticism as it is a self-report (with self-serving biases and protected interests), social and interpersonal protocols and trust’s socio-emotional and status-laden characteristics can make it very difficult to reject such an appeal. A rejection disputes the accuracy of the would-be trust recipient’s self-assessed confidence in their ability, benevolence and integrity. A rejection inevitably casts negative aspersions on the trust recipient (especially their status) in a way that could be damaging to the relationship, particularly if declining the request is carried out in public view. Moreover, the implied slight may require a justification, which might be socially awkward to proffer. Thus, a trust recipient can, whether intentionally or unintentionally, use the language and dynamics of ‘trust’ to solicit trust from a possibly reluctant trust giver. This feature distinguishes the scenario from the misplaced trust that results from an error of judgement on Ben’s part – although to bestow trust too readily may be seen as incompetence by others, heightening the anxiety around the ‘trust/don’t trust’ decision. Trust can also become a poisoned chalice in this scenario if Anthony’s exhortations as to his trustworthiness persuade Ben to develop the ‘confident positive expectations’ necessary for the trust decision despite his initial misgivings.

Examples in organisations include peers being asked to collaborate with colleagues they believe to be unreliable or inadequate. The Anthony in our scenario
might be a salesperson with a reputation for dull or incoherent presentations, but who insists on being given the chance to redeem himself by delivering his team’s pitch to a major client. Another example is having to trust management statements on future job losses (cf. the Hardy et al., 1998, case study mentioned earlier) or claims of sound corporate governance practice and appropriate risk assessment (for example, Dick Fuld of Lehman Brothers and Andy Hornby of HBoS prior to each firm’s collapse). Fox observed this scenario in manager-subordinate relationships; ‘we’ve got to trust them’ means in fact ‘we don’t trust them but feel constrained to submit to their discretion’ (Fox, 1974: 95). Graebner (2009) found that many entrepreneurs commit to contractual agreements with investors against their better judgement, because of deliberate trust-inducing deceptions.

In the trust literature, much emphasis has been placed on the downside of inappropriate trust for the trust-giver (Gargiulo and Ertug, 2006; McEvily et al., 2003). However, the solicitation of trust is not without a price for the recipient (Anthony in our scenario). Issuing an appeal to be trusted increases the trust recipient’s vulnerability to a failure to deliver on expectations. Anthony would have no room for error in making his pitch. Employees who consent to place trust in their line management, for example, form a clear expectation that those managers will protect the employees’ interests. Failure to meet these expectations is likely have a detrimental effect on levels of staff trust, and consequently on commitment, motivation and morale. From their perspective this may be an unwelcome implication for that management group’s obligations. The scenario is still of misplaced (inappropriate) trust because the recipient sees it as inappropriate - they know they can’t, or don’t wish to, fulfil the implied obligations.

This leads us to our second poisoned chalice scenario, in which the roles are reversed to consider the effect of unwanted trust for the trust recipient. There are a number of reasons why somebody doesn’t want to be trusted; it may be because of their extended reciprocation wariness (Eisenberger, Cotterell and Marvel, 1987) or
they may view any trust placed in them as inappropriate. Wary employees have been found to react to social exchanges with less affective commitment and trust, and with higher turnover intentions than less wary employees (Shore et al., 2009). However, the potential for the ‘poison chalice’ lies in scenarios when the recipients view trust as 
unwelcome, because (s)he fears that the trust-induced obligation cannot be repaid. This creates a scenario in which the trust recipient does not wish to be trusted because they doubt that they can fulfil the obligation or they do not wish to do so, but they find the offer difficult to refuse.

Scenario 2: Unwelcome trust – in which Andrew wishes to place trust in Brenda but Brenda doesn’t want to be trusted.

In this scenario, the trust recipient (Brenda) is informed that she is being trusted (by Andrew). The trust decision has been taken by Andrew, but prior to the risk-taking act Brenda faces a dilemma: she may prefer to decline Andrew’s trust in her because it communicates expectations, and possibly obligations, relating to future behaviour that may be problematic, but she may feel unable to do so. This reluctance to be trusted may occur when trustees, such as Brenda, doubt their capability to fulfil the obligations implied by the trust (a low self-report on their own ‘ability’, such as weak self-efficacy beliefs) or because the pressures of the obligation may be intolerable (for example they may believe that there could be damage to their mental and/or physical well-being if they attempt to fulfil the trust-related expectations), or may run counter to their interests (Flores and Solomon, 1998) or personal ethics. In this sense, to be trusted constrains the choices of the trust recipient and can be unwelcome. This particular exchange dilemma may be further exacerbated, however, because, as we have seen, the act of trusting implies a favourable, even admiring assessment of the trust recipient that one would like to vindicate for identity (social-identity cues; socio-emotional) reasons. This is particularly the case if the giver is a significant valued partner such as a superior or esteemed colleague (trust’s particularism).
Andrew’s intention to trust Brenda may be positive and genuine, but when trust is neither sought nor welcome, it can become a negative experience – a poisoned chalice – for the recipient. Unsolicited trust presented in such a way as to be difficult to reject may be used to inspire and motivate but, given the status and kudos benefits claimed for being trusted, this dark side scenario suggests that the trust-giver could also use ‘trust’ in such a way as to create a trap of obligation.

An organisational example is when managers ‘empower’ under-skilled and/or unmotivated employees (Allen et al., 1997). For the under-skilled trust recipient, a lack of self-efficacy may mean that empowerment pitches them into decision-making beyond her/his capability. Additionally, not all employees may want to be empowered (i.e. trusted); in many cases ‘empowerment’ is perceived by recipients as a means of increasing their workload and responsibilities without a commensurate pay rise (Mabey and Skinner, 1998). Another might be being designated as ‘top talent’, if the concomitant expectations for performance and additional responsibilities are unattractive and burdensome. A further example would be a supplier given a near-impossible deadline by a valued customer for whom a refusal to comply would be highly damaging to the relationship. If trust is unwanted but accepted, it may lead to work intensification and stress (Green, 2004) and forced errors, with knock-on repercussions for the trust that was unwanted in the first place.

*Scenario 3: Trust lock-in in long-standing relationships – Alan has placed trust in Brian and continues to trust him but Brian no longer wishes this to be the case.*

Our third scenario is an extension of the second, and illustrates the use of long-standing social exchanges and deep bonds of indebtedness (one party’s power over another) to create a ‘trust trap’. The same sequence and logic applies in this scenario as in the previous one, but the length of the relationship and the shadow of past exchanges intensify the dilemma. Brian’s trust relationship with Alan may have begun positively but, over time, being trusted has ceased to be beneficial for Brian. He may no longer be able or willing to fulfil what is expected of him because of incapacities or
competing demands. He is now a reluctant trust recipient. As such, any pre-existing moral or contractual joint agreement may be viewed as a constraint if the terms of the deal (and the obligations implied by them) threaten his interests, and the agreement is too inflexible to accommodate revisions to the rules of engagement.

SET suggests that the more powerful the social bond (the deeper the investment in trust and the resulting indebtedness), the harder it is to reject it, or to act contrary to its implied obligations (Coyle-Shapiro and Conway, 2004). This accounts for the rigidities and lock-in effects which may arise in long-standing and successful relationships. In trust scenarios, when its particularistic quality dominates the decision, there is a danger of an ‘amplification of reciprocity’ that leads to ‘structural arthritis’ in relations (Gargiulo and Benassi, 2000: 185).

An organisational example is the fallout from a failed partnership agreement between a management team and a trade union. Brian, as a union representative, may have reached a ‘line in the sand’ beyond which he cannot reconcile himself to a painful concession, such as job losses or a pay freeze, ‘partnership’ obligations notwithstanding (see Gall, 2001). The situation of a supplier firm’s long-standing contract with a dominant and unreasonable customer (Cox, 2001) is another example, as is a failing alliance (Adobor, 2006). In these scenarios, trust’s socio-emotional and particularistic qualities – once prized and cherished – can become problematic and even intolerable.

Again, the predicament of reneging on a long history of trust is interpersonally awkward, and is likely to come as a shock. The party still trusting might not easily decipher the cues of the relationship-retreating party. When the retreating party’s change of mind becomes apparent it is likely to be perceived as a breach or, even more strongly, as a betrayal, resulting in emotionally-laden and identity-threatening negative experiences. Hence, a long history of productive trust exchange can become a burden when trust is no longer attractive. The awkward social challenge of managing withdrawal from a relationship can lead to ‘commitment escalation’
(McAllister, 1997: 101) on the part of the reluctant trustor. The long-standing ‘trust’
obligation may seem the lesser of the two predicaments.

These scenarios have highlighted how both the social exchange and trust
literatures assume parties have a vested interest in fulfilling obligations and
furnishing inducements, to demonstrate their trustworthiness and nurture trust. Yet
this is not necessarily the case. In his original work Blau saw that social exchange
principles and dynamics can create pressure to discharge unwelcome obligations. He
noted, crucially, that failure to comply with obligations can be ‘disadvantageous’ by
leading to poorer relations, but also reputational damage and guilt for the dissenting
party (Blau, 1964:.97). Our argument is that this dynamic, based on the norm of
reciprocity, is present in trust, and can be manipulated to be coercive, creating bonds
of indebtedness (Coyle-Shapiro and Conway, 2004), trapping people in unproductive
but inescapable relationships (Gargiulo and Benassi, 2000). What might distinguish a
positive chain of obligations from a pernicious one is motive.

The importance of motives. Griffin and O’Leary-Kelly (2004: 5) suggest that the same
behaviour can be either functional or dysfunctional, ‘depending on intent, motive,
context and consequences’. As noted earlier the gift-giving literature highlights the
importance of motives more strongly than the SET literature. Extending this thinking
to trust highlights further possible poison chalice scenarios which arise from the
motivation of one or both parties in the relationship. People may baulk at suggestions
of self-interested intentions being behind trusting interactions, but we have already
seen in scenario 1 how having to trust someone creates an unbalanced power
relationship and that those with power may use appeals to be trusted as a means of
influencing weaker counterparts’ behaviour (Brandts and Charness, 2003), and/or to
avoid close scrutiny of their actions. We have also seen, with scenarios 2 and 3, how
being trusted can entrap a wary partner, effecting a similar asymmetric reciprocity
that, in those two scenarios, underscores the trust-giver’s supremacy.
This potential for an asymmetry of power in trust relationships, especially those crossing a hierarchical divide (such as employees and their leaders; buyers and suppliers), has been underplayed in the literature. However, it is here where certain characteristics of trust may become most apparent and most salient. Put simply, trust's symbolic and emotional connotations have the potential to be exploited for political manoeuvre, or manipulation, because, as we have argued, to suspect or to refuse trust might create all kinds of negative consequences. The identity of the focal party – whether as trustee or trustor (i.e. trust's particularism) – may be a decisive factor in facilitating a manipulative ‘trusting’ power-play and the intentions behind the trust exchange are critical for the experience of trust.

**Scenario 4: Withdrawn trust – Anouk decides not to continue to trust Bruno.**

In this scenario, a sequence of decisions to trust has come to a deliberate end. Anouk’s withdrawal of trust casts powerful negative aspersions on the trustworthiness of the now spurned recipient, Bruno, whether in relation to his ability, benevolence or integrity, or a combination of all three. Experiencing a trust withdrawal might be expected to damage Bruno’s self-image and reputation, and even jeopardise his standing in social networks. His distress may come from shame, embarrassment and even guilt from being no longer trusted. Losing someone’s trust may, therefore, render it a poisoned chalice in retrospect. Examples from organisational life include being pulled from an important and high-profile project, being passed over for a promotion, or a public sacking. The stigmatisation felt as a result of ostentatious new monitoring mechanisms also represents this scenario (for example daily appraisals, performance league tables and revised codes of ethical conduct for professionals – see O’Neill, 2002).

Withdrawals of trust may be justifiable, even on explicit trust grounds (due to diminished confidence in a focal party’s trustworthiness) but it is a manifestation of trust’s dark side when threat of its removal is used as a sanction and a control mechanism (Mabey and Skinner, 1998), to secure the trust recipient’s
cooperation/submission. An example would be Anouk indicating to Bruno that his continued membership of a prestigious team, or even his promotion (both suggestive of her ‘trust’ in him) is dependant upon Bruno carrying out certain tasks. As with scenarios 1 to 3, this scenario points again to the intriguing potential for trust to serve as a control. Moreover, if Bruno considers the tasks demanded of him to be untrustworthy acts that he must nevertheless carry out, his discomfort will likely be heightened as he now faces scenario 2 (not wanting to be trusted), as well.

The distress for the no longer trusted party can be especially problematic due to the salience of trust’s particularism and status-laden characteristics for the trust recipient: if the relationship is important, the consequences of trust’s withdrawal will be especially serious. In addition, trust withdrawal conducted publicly is likely to have a more deleterious effect; a private withdrawal can be expected to be less traumatic than a public humiliation (trust as a ‘status-laden’ commodity).

Scenario 4 has echoed the valence of being trusted from scenarios 2 and 3. Our final scenario looks at the distress that occurs when trust is promised, but not delivered. In this scenario trust is insincere: trust is promised, perhaps as an incentive, but it is only given in word but not in deed.

**Scenario 5: Insincere trust – Alison expresses trust in Bethany but this is not supported by her actions.**

In this scenario, although ‘trust’ is articulated in the form of a professed intent by the trust-giver, it is not demonstrated by the risk-taking that is trust’s most convincing manifestation. The promised trust never arrives. Or, worse from Bethany’s point of view, Alison’s stated intent is accompanied by trust’s opposite, close monitoring and tight control: ‘If you [truly] trust a partner, you do not need to verify [i.e. check up on them]. Doing so would be the clearest indication that you do not trust [them]’ (Schoorman et al., 2007: 350). The prospect of being trusted, particularly if it is something sought after, and especially from a valued counterpart (trust’s status-laden value), means that its denial in practice renders that initially declared
trust a poisoned chalice, because of the implicit aspersions that its denial casts on one’s character, or competence and the frustrations which are inherent in the situation.

Micro-management is the most obvious, and commonplace, example (Wright, 2000), such as an autonomous project leader (Bethany) to whom responsibility has supposedly been delegated who must nevertheless report back to her manager (Alison) prior to any major decision. Bethany’s autonomy is a chimera, as she has been given no real power; she is not really trusted. Another example would be a superficial empowerment scheme that tolerates employee discretion in trivialities but debars their input into more substantial workplace and strategic concerns. The trust implied in empowerment is negated by the subsequent lack of risk-taking. A common tension point for trust comes with organisations’ legal and compliance procedures (O’Neill, 2002): an employer may say it trusts its employees, but the onerous stipulations of its compliance regime suggest otherwise.

Based on these last two scenarios, and given trust’s socio-emotional and identity-communicating functions, the refusal to ‘walk the talk, to follow-up on the promise of trust, is likely to invoke strong negative reactions on the side of the (non-)trust recipient. Indeed, Kickul and co-authors (2001) have shown that employees react most strongly when they are promised autonomy and empowerment, but perceive that these promises are not enacted. Brockner and Siegel (1996) found that individuals with high organisational commitment react more harshly to their employer’s broken promises than those with low commitment, because the former perceive the unfairness as an identity-threat.

Yet, we also see this scenario being problematic for the trustor (Alison). Taking the empowerment example, the fear of poor employee input has been cited as one reason why line managers find empowerment of their direct reports such a threat (Spreitzer and Mishra, 1999). Trusting her subordinates may be perceived by Alison as an unacceptable risk and even a challenge to her self-image as an
experienced and qualified manager responsible for taking the organisation’s decisions. She may view trusting staff as a threat to her managerial status and even her job security (Batt, 2004). However, if Alison says that she has confidence in Bethany but fails to act upon it, the disconnect between her word and deed damages Alison’s integrity (Simons, 2002).

This particular scenario goes to the heart of the debate about whether trust and risk-taking acts are distinct, though related, phenomena (as with Mayer et al., 1995) or whether risk-taking is the only credible demonstration of trust and hence its definitional realisation (Schoorman, et al., 2007). This scenario – of insincere trust being unsettling, even demoralising for the recipient – neatly encapsulates the argument in favour of the latter. Trust cannot only be viewed as a ‘psychological state’, as per Rousseau and colleagues’ (1998) widely used definition. The act is real trust, not the stated willingness to trust.

This scenario also highlights the relationship between trust and forms of control such as monitoring, or contractual protections (Weibel, 2007). While the scenario might appear to support the substitution thesis of an inverse relationship between trust and control (the institution of governance mechanisms is seen as informed by suspicion and hence is inimical to trust), we believe it suggests instead that controls which obviate any risk for the trust-giver may be unwelcome, and may render talk of ‘trust’ hypocritical. However, some forms of control may be acceptable (the complementarity thesis). Indeed, the bulk of the available evidence suggests that trust and control can work as complements rather than substitutes (Costa and Bijlsma-Frankema, 2007): you can have both (trust based on controls, or acceptance of controls because of the presence of trust). Bethany may accept some monitoring of her work, if it is reasonable; indeed, she may see checks and balances as indicative of Alison’s competence and of her genuine interest and care. But she is likely to resent punishment-oriented, or suspicious, control when she is supposedly trusted. This suggests that the content of the controls, the intent behind the controls,
and the trustee’s consent to the controls, are all decisive moderators of this relationship (Das and Teng, 2001).

The Research Agenda.

If we understand the process of trust to be one which is inherently relational and involves a series of exchanges then the ideas found in the social exchange and gift-giving theory offer a useful lens through which to extend our understanding of its nature. These literatures explore the complexities of a dark side to social exchanges and it seems unlikely to us that trust would be exempt from this aspect of human interaction. Accepting this potential allows us to challenge the apparent truism that trust is always, in and of itself, something which is good and desirable. We believe that this truism limits our ability to understand the full nature and complexities of trust. In our analysis of trust’s dark side we have identified some of the conditions under which trust is a welcome and positive experience, and more importantly for our purposes, when it has the potential to be unwelcome and negative. It is widely assumed in the trust literature that evidence of Mayer et al.’s (1995) three trustworthiness attributes (ability, benevolence, integrity) or Whitener et al.’s (1998) five categories of trustworthy behaviour (consistency, integrity, sharing and delegation of control, communication, and benevolence) are sufficient to foster ‘real’ trust. The conventional view of trust’s ‘dark side’ follows from this: it only arises when this assessment is misjudged. Our essential argument is that the default assumptions in the trust literature of benign motives, welcome obligations and productive reciprocity, do not cover all trusting scenarios. We do not deny that purely positive trust experiences are possible but there are also those trust experiences that can result in an unwelcome burden and which may stem from less than benign intent. In our illustrative trust scenarios we have suggested ways in which trusting interactions could generate obligations that are unwelcome for the focal party, harbouring potentially negative consequences. These are trust’s ‘poisoned chalice’ scenarios.
and the possibility of such experiences highlights the need for new directions in trust research if we are to fully understand this very complex aspect of human interaction within organisations.

We need to understand precisely what is being exchanged when parties trust each other, how the process unfolds and how cues are recognised and interpreted. SET suggests that trust is symbolic, status-laden, socio-emotional, intangible and particularistic, and we have argued that these characteristics embody trust’s innate value but therein also lies the potential for trust to be a poison chalice. A dominant assumption about trust is that parties must share mutually positive outcomes in trust-based relations. Though most trusting interactions are likely to be honest and sincere, our analysis has shown that this assumption masks the full range of interests and intentions involved. The social pressures innate to the trust experience, as well as the difficulties in assessing others’ trustworthiness accurately, creates the possibility that the process of trust could be used for deception. This echoes a conclusion reached by Brenkert (1998b: p. 281): ‘the tendency... to equate trust and morality must be rejected’. Neither being trusted nor trusting is always benign. Nor does trust always leads to positive consequences for all parties involved. Returning to Blau’s two motives for social exchange, our scenarios illustrate either misjudged efforts to realise a positive relationship (a ‘friendship in Blau’s terms) which could conform to the conventional depiction of the dark side or deliberate attempts to achieve super-ordination using the dynamics and/or the language of trust.

This suggests a number of fruitful areas for research including how people actually process cues of trustworthiness; how they understand and interpret both trust related language and actions and the importance of the context in which these take place. Repeated over time, trust exchanges create a cycle of obligations to be fulfilled (Whitener et al., 1998). The gradual expansion of exchanges that have intrinsic and/or extrinsic appeal for both parties becomes motivationally self-enforcing, binding trust giver and trust recipient together (Pettit, 1995). Yet, for trust
to develop, both parties must abide by certain rules which serve as the ‘guidelines of exchange processes’ (Cropanzano and Mitchell, 2005: 875). However, these guidelines are vague and ill-defined for general exchange processes, and the problem with trust is that its rules are not at all obvious. There is little agreement on the established rules and criterion that govern such social exchange interactions. It is likely that economic logic cannot be applied to social exchanges and that there are different rules for the exchange of different resources; love is exchanged differently to money. Nor do we understand the time-frame for trust’s reciprocity other than to know it is not uniform but is contingent on the context and the relationship, indeed some argue that ‘for trust to develop the recipient should not reciprocate hastily’ (Shore et al, 2009: 290) but is this always true or only true in certain circumstances? There is a need to better understand the ‘rules’ that individuals and social groups operate, whether consciously or unconsciously, in the context of trust exchanges. The SET and gift giving literature note the role of social expectation and the additional pressures associated with exchanges which occur in public. Intuitively we would expect the same to be true of trust exchanges. The power balance within the employment relationship is another important contextual factor which has received limited consideration in relation to trust but is one which is of particular interest not only in exploration of the factors which lead to an assessment of trustworthiness but also the potential to contribute to the creation of a poison chalice experience.

The second area for research relates to the motives which lie beneath these exchanges and the scope for manipulation. While the widespread agreement in the literature is that trust can neither be mandated nor bought (Gambetta, 1988), we question whether this assumption is correct. Trust based on positive well-intentioned confidence in another has the potential to result in a negative experience, perhaps because it is misplaced or not desired by the recipient, but we also believe that there is the potential for trust to be consciously used to manipulate both individuals and situations. Manipulation is defined as a compilation of three features: some sort of
advantage accrues to the manipulator (Porter et al., 1981; Schminke, 1998); a deception is used such that those being influenced are not consciously aware of the attempt to control them and, finally, the manipulator’s motive is to change the other party’s behaviour to achieve the manipulator’s own objectives (Green & Pawlak, 1983; Wrong, 1988). Manipulation often comes in the form of tactics related to the use and availability of information and creation of favourable images (Allen et al., 1979). Deception is typically accomplished by a controlled distortion of the perception of reality (Porter et al., 1981). Social exchange resources “cannot be bartered lest their value as genuine feelings or judgements be compromised” (Blau, 1964: p.96). So deceivers have to rely on the language of trust to hide their intent however, the discourse of trust – how the word is used in social interactions – is an under-explored and potentially significant research agenda. Invocations of trust could be used as linguistic ‘smoke and mirrors’ to create the unwelcome obligations or inferences that compel another to act against their own interests in the pursuit of trust. In other words, the very language of trust may contain within it a sinister potential as deliberately engineered performative acts (see, for example, Hodgson, 2005). Several of the illustrative scenarios depict controlled distortions of the notion of trust, whether by constraining a reluctant trust recipient’s options within trust parameters set for the trust donor’s benefit (scenarios 1, 2, 5), or by compelling a reluctant trust recipient to trust despite their better judgement (scenario 3). We stress, however, that this is not the case for all scenarios, some may be unintentional. So, while each of the poisoned chalice trust exchanges are nominally reciprocal in that trust is being exchanged, what distinguishes genuine reciprocity and genuine trust might be the motives and intentions behind the exchange.

Our understanding of trust as we have explained it here includes the fundamental importance of the risk-taking act for trust. This tends to be considered separate to the trust process itself (Mayer et al., 1995: p.724), but in our opinion is its natural manifestation; (Schoorman et al., 2007). As illustrated in the fifth scenario
trustworthiness beliefs and a willingness to be vulnerable may be superficial, even hypocrical, without the act of taking a risk. The intention to trust may not be carried through, and hence not be experienced as trust; worse, intent alone can be re-cast as distrust without an accompanying risk-taking act. In short, understanding trust as the three-stage process we believe that it is requires that the risk-tasking act receives more attention. This has major implications for the design of empirical studies. Measures of trust should incorporate risk-taking acts (Dietz & den Hartog, 2006), whether by proxy (e.g. manifest in a decision), or from the creation of a new measure of trust as a behaviour, similar to that for organisational citizenship behaviour (OCB).

**Conclusion.**

It is our belief that it is time to challenge the rather glib, ‘monochromatic’ optimism (Sievers, 2003: 360) that dominates most literatures on trust such that trust is widely considered to be inherently a good thing, almost untouchably so. In this paper we have explained why this state of affairs is limiting and have called for a wider exploration of dark side behaviours in the realm of what Sievers called ‘one of the few white spots [i.e. untarnished – sic] left on the organisational map’ (2003: 358). In doing so our intention is to provoke both debate and challenge and to begin the development of a research agenda which seeks to better understand the dark potential we see in the complexities of trust.
References


toward emotional well-being’. *Psychological Science*, 13, 172-175.

and Faber.

York: Free Press.

Gall, G (2001). ‘From adversarialism to partnership? Trade unionism and industrial
relations in the banking sector in the UK’. *Employee Relations*, 23, 353-375

Basil Blackwell.

structural holes, and the adaptation of social capital’. *Organization Science*,
11 (2), 183-196.


Jones, C (2003). 'As if business ethics were possible, 'within such limits...'

*Organization*, 10 (2), 223-248


http://www.bbc.co.uk/radio4/reith2002/


