This is a repository copy of The impact of self-affirmation on health-related cognition and health behaviour: Issues and prospects.

White Rose Research Online URL for this paper: http://eprints.whiterose.ac.uk/43100/

Article:
doi: 10.1111/j.1751-9004.2010.00270.x

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher’s website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
The impact of self-affirmation on health-related cognition and health behaviour: Issues and prospects.

Peter R. Harris and Tracy Epton
Department of Psychology, University of Sheffield, UK

p.harris@sheffield.ac.uk
t.epton@sheffield.ac.uk

This is a pre-copy edited and proofs version of the paper that appears in its final form as:


The published paper is the definitive final version and differs in a number of respects from this version.
Abstract

In a recent paper (Harris & Epton, 2009) we reviewed the evidence showing that self-affirming – the act of reflecting upon cherished values or attributes – can reduce resistance to health-risk information. In this companion paper we extend the discussion of issues arising from that review and describe key questions for future research. Overall, we regard the picture emerging from this nascent literature as encouraging. Nevertheless, more needs to be discovered about how self-affirming achieves its effects and their limits. Despite lowering an important barrier to health behaviour change by reducing message resistance, there is currently only limited evidence that self-affirming changes subsequent health behaviour. We consider why. We also discuss issues to address in interventions involving self-affirmation and examine evidence that self-affirming alters relationships between variables. There is also scope for extending the range of samples, health information, and health behaviours examined and for assessing more spontaneous self-affirmation.
Most people prefer information that supports rather than challenges their important beliefs. This preference has been shown for beliefs of many kinds (Sherman & Cohen, 2006). In the case of health beliefs, a preference for congenial information may lead people to reject potentially vital information by, for example, downplaying the importance of messages about health risks, or prompting them to regard such information as personally irrelevant. Consequently, those for whom the information is most relevant may be the least persuaded about the long-term dangers of their lifestyle and the need to change their behaviour (Good & Abraham, 2007).

There is evidence that self-affirming – the act of reflecting upon one’s important values or cherished attributes – can reduce such resistance to uncongenial information, including resistance to unwelcome but important health-risk information (Harris & Epton, 2009). Self-affirming can promote greater acceptance of health-risk information and reduce message derogation in the target audience; self-affirmed participants often also show greater interest subsequently in changing their health behaviour. In this paper we consider some of the issues arising from the previous review and the questions we think are in need of further examination. In preparing it, we have assumed readers are familiar with its companion (Harris & Epton, 2009).

**Self-Affirmation Theory**

Self-affirming is the process of reflecting upon one’s cherished values, actions or attributes. It functions to restore or maintain a person’s sense of who they are and what they stand for when they experience threats to their identity. The theory proposes that people are strongly motivated to maintain their sense of being rational, decent, sensible people (encapsulated in the theory as their sense of “self-integrity”),
Steele, 1988, p. 262). So, when they feel their self-integrity is threatened they take steps to protect it.

From this perspective, one reason why people resist messages warning them about risks they are taking with their future health is that such warnings threaten their self-integrity. Resistance – “a motivated state in which the goal is to withstand the effects of a persuasive communication” (Jacks & O’Brien, 2004, p. 236) – is one strategy that preserves self-integrity; if the message can be undermined and rejected, then it poses no threat. However, intriguingly, self-affirming has the potential also to reduce such “defensive” resistance to threat. Self-affirmation theory proposes that people are motivated to maintain their overall or global sense of self-integrity. Consequently, a separate reminder of something self-affirming – even something completely unconnected to the threat – can be sufficient to buffer against that threat. Thus, self-affirming can promote more objective appraisal of otherwise threatening information. For example, recalling recent occasions in which you acted in accordance with an important value (e.g., kindness or honesty) can bolster self-integrity; so, if shortly afterwards you face a message about the risks of continuing to smoke, you feel able to process that message with a more open mind (because you are no longer so concerned that it will damage your self-integrity).

This prediction – that self-affirming can promote more objective appraisal of otherwise threatening information – has been tested (and found to be very largely supported) in the literature reviewed by Harris and Epton (2009). In a typical experiment participants are required to self-affirm (e.g., by writing about an important value) before being asked to examine relevant health-risk information (e.g., about the risks of excess alcohol consumption). Their responses to this information (e.g., on
measures of message acceptance, affect, and intentions; see Harris & Epton, Table 1) are then compared with those of a control group.

**Some Benefits of the Self-Affirmation Theory Approach**

It is important to understand resistance to health-risk information for both theoretical and practical reasons. Theoretically, we need to improve our understanding of the mechanisms underlying resistance. Practically, resistance represents an important barrier to health behaviour change; if people do not accept that a message is relevant to them, they are less likely to attempt to change their behaviour (Weinstein 1988).

Research on self-affirmation has the potential to contribute to both theoretical and practical understanding. Self-affirmation theory emphasises the role of the self-system in understanding why people are resistant to unwelcome information. This enhances existing perspectives on resistance to health-risk information, which tend to emphasise physical threats rather than threats to the self (though see Das, de Wit, & Stroebe, 2003, for an exception). Self-affirmation theory also suggests a number of relatively simple ways of reducing message resistance that may have potential for development as applied techniques for use in interventions; interventions that are easily implemented are in great demand. Moreover, though there may be problems in using self-affirmation manipulations in applied settings – as we discuss later – they have one major advantage over most alternatives: they do not require changes to persuasive materials. Instead, it is proposed, self-affirming changes the way people approach and respond to information, rendering them more open-minded and prepared to accept strong and persuasive information telling them things they would prefer were not true. As well as practical advantages – it allows existing materials and information to be used – this also has ethical advantages, as it is not a technique for increasing persuasion. Instead, self-affirming affords more objective appraisal of
existing information – allowing it to “speak for itself” – and therefore potentially contributes to more genuinely informed decision-making. That is, self-affirmation reduces message resistance, but whether that results in message acceptance depends on the quality of the information.

**How does self-affirming reduce resistance to relevant health-risk information?**

The principal evidence that self-affirmation reduces resistance to health-risk information is that it typically increases message acceptance in at-risk groups (Harris & Epton, 2009). However, there are limitations to the existing studies that mean many questions concerning the boundaries of the effects, as well as about the processes involved, remain unanswered. Below we consider some of the questions we think it would be useful to address in the next phase of research.

It would be helpful to know more about where in the stages of message processing self-affirmation makes a difference and how it affects the strategies people use to resist persuasion (Jacks & Cameron, 2003). For example, Blumberg (2000) describes four defensive coping strategies that people use to resist: attention avoidance (indiscriminately avoiding all relevant messages), blunting (avoiding the threatening elements of the message), suppression (trying not to think about or elaborate on the self-relevance of the information), and counter-argumentation (biased assessment and active refutation of elements of the message). Frameworks such as this could be used to establish where self-affirmation begins to change people’s cognitive and affective response to threatening health-risk information and whether and how it shapes those responses at different stages. There is some evidence that self-affirming may reduce blunting. For example, in Klein and Harris (2009) self-affirmation enhanced attentional bias towards threatening words taken from the message in at-risk participants. Similarly, in Van Koningsbruggen, Das and Roskos-Ewoldsen (2009)
self-affirmation increased the accessibility of threat-related cognitions in the target audience.

Does self-affirming reduce attention avoidance? Studies to date have required that participants expose themselves to the threat, so we cannot tell. Indeed, apart from one early study (Reed & Aspinwall, 1998), no study has allowed participants to navigate their own way through the information. What happens when people are free to structure their own passage through material or allowed to avoid parts of it? Do at-risk, self-affirmed participants differ in how they search for and select material on the Internet, for example? The Internet contains information that both threatens and reassures and there is evidence that people may be more naturally drawn to the latter (Dutta-Bergman, 2004; Joinson & Banyard, 2003; Sillence, Briggs, Harris & Fishwick, 2007); it would be reasonable to predict that self-affirmation changes this in favour of more balanced data-gathering, but this has yet to be tested.

Does self-affirming reduce suppression or counter-argumentation in response to health-risk information? Little systematic attention has been paid to these questions. Indeed, little attention has been paid so far to process questions in this literature (Harris & Epton, 2009). While self-affirming is hypothesised to promote greater open-mindedness, does this typically result in more or less thoughtful processing? Does it alter processing goal (Chaiken, Gina-Sorolla, & Chen, 1996)? Self-affirmation has been shown to increase sensitivity to argument strength in non-health domains (Correll, Spencer, & Zanna, 2004), suggesting it does result in more central route (thoughtful) processing (Petty & Cacioppo, 1986). However, self-affirming before a persuasive message appears also to boost judgmental confidence; where the message is low in threat, such confidence leads to less thoughtful message processing (Brinol, Petty, Gallardo & DeMarree, 2007). There is considerable scope for research
investigating when and in what ways self-affirming affects the processing of health-risk information and people’s thoughts about their thought processes (or meta-cognitive judgments), such as judgmental confidence.

When and how does self-affirming moderate the affective response to health-risk information? Surprisingly few studies have examined these questions, especially given the central role that negative affect, such as fear, is believed to play in both threat minimisation and message responsiveness (e.g., Witte & Allen, 2000). Likewise, few studies have tested how self-affirming affects self-related affect. Where studies have examined affective responses, few have used implicit measures of affect, even though some theoretical explanations of the effects of self-affirmation rely on implicit mood as a mediator (e.g., Tesser, 2000). There is also a need for further tests of the intriguing idea that self-affirming promotes open-mindedness by enhancing other-directed positive emotions, such as love (Crocker, Niiya & Mischkowski, 2008).

In choosing dependent measures, more explicit attention might usefully be paid to the literature on how to identify and measure threat minimisation, defensiveness and responsiveness. For example, following their meta-analysis, Good and Abraham (2007) suggested that message acceptance, perceived severity and susceptibility are reliable measures of defensiveness, but also that some measures (e.g., perceived susceptibility) appear less sensitive to self-affirmation manipulations than others (e.g., message agreement). Which measures are included, therefore, may influence what is found. In turn, Rothman and Salovey (2007) point to the need for multiple indicators that assess responsiveness to the information as well as threat minimisation, as these may occur in parallel (Leventhal, 1970). Self-affirming appears to promote responsiveness over minimisation in at-risk groups, and thus to alter the balance
The impact of self-affirmation (Harris & Epton, 2009). We need to understand more about the implications of changing this balance, such as its consequences for successful health behaviour change. This is a key issue to explore in future research.

In testing how self-affirming affects responses to health-risk information researchers could use more diverse samples and types of information. In particular, the literature would benefit from testing how self-affirming affects responding in samples with genuine health issues, such as those coping with chronic conditions. Health-risk information also comes in many forms other than the one-sided, verbal, non-tailored, persuasive messages that currently predominate. Even within the persuasive message paradigm there is scope to broaden the focus from health compromising behaviours (like smoking) to health-promoting behaviours (such as taking exercise) and from preventive to detection behaviours (such as screening for disease). Indeed, there is a disjunction between the self-affirmation literature and that on defensiveness to health information, which has explored detection behaviours in some detail (e.g., Croyle, Sun, & Hart, 1997). Detection behaviours are perceived as riskier and respond better to loss-framed appeals (Rothman & Salovey, 2007). In extending the range of behaviours targeted researchers might test whether framing moderates the effects of self-affirming, particularly as a function of behaviour type. For example, it may be that self-affirmed participants are less sensitive to the framing of an outcome as a loss or a gain than their non-affirmed counterparts.

Messages have also tended to target health threats that are both temporally remote and statistically unlikely for the samples of young people involved: investigating more proximal health issues in such samples would be useful. Moreover, with only a few notable exceptions (Epton & Harris, 2008; Fry & Prentice-Dunn, 2005) relatively little attention appears to have been paid to the structure and content of the message,
which could be developed using more theory-based approaches (e.g., Maibach & Parrott, 1995; Witte, Meyer & Martell, 2001).

Self-affirmation induced open-mindedness

What happens to people when they self-affirm and why does this make them more open-minded? In Harris and Epton (2009) we discussed possible mediators, including several that show promise, such as boosts to self-certainty and to other-directed feelings; However, currently there is no generally accepted mediator.

More broadly, Sherman and Hartson (in press) have recently proposed a three-stage model of how self-affirmation promotes more open-minded appraisal of unwelcome information. According to this model, self-affirming boosts self-resources, defined as those “objects, personal characteristics, conditions, or energies that are valued by the individual or that serve as a means for attainment of these objects, personal characteristics, and energies” (Hobfoll, 1989, p. 516). This boost to self-resources gives people the energy and scope to confront the threat rather than rationalising it. In particular, having self-affirmed they feel more secure in their self-worth and thus less concerned about its self-evaluative implications. Indeed, according to Sherman and Hartson, self-affirmation actually “uncouples” the threat from the self so that the threatened domain no longer contributes as much, if at all, to self-evaluation. Recent findings are consistent with this model. For example, Schmeichel and Vohs (2009) have shown that self-affirmation counteracts the effects of depleted self-regulatory resources and improves self-control by promoting higher levels of mental construal (see also Wakslak & Trope, 2009).

Although Sherman and Hartson are careful to point out that they do not presume that their model describes the process underlying all self-affirmation effects, it provides a compelling synthesis of the current literature. In part derived from the
literature on self-affirmation and health-risk information, it may provide a useful framework for future analyses of how self-affirming affects health risk processing.

Nevertheless, much remains to be discovered about the processes by which self-affirmation enables people to become more open-minded about information describing threats to their health and the limitations on these positive effects.

**Moderators of the effects of self-affirming on responsiveness to health-risk information**

Several important leads have emerged about the boundaries of the positive effects of self-affirming on responsiveness to relevant health-risk information. Other possible moderators have been suggested but await empirical scrutiny.

Findings to date suggest that self-affirming not only appears to reduce message resistance among more at-risk groups, but that the effects can be most pronounced in these groups (Harris & Epton, 2009). However, Van Koningsbruggen (2009) has argued that the benefits are limited to moderate threat levels and that self-affirming promotes less rather than more objectivity when threat is high or low. This claim is obviously of both theoretical and applied importance but awaits empirical scrutiny. In Klein and Harris (2009) changes in attentional bias after self-affirming were indeed limited to moderately heavy drinkers, but self-affirming at higher risk levels did not induce bias away from threat. Indeed, self-affirming has been hypothesised or found to have its biggest effects (e.g., Harris & Napper, 2005), no effects (e.g., Klein & Harris, 2009) or rebound effects (e.g., van Koningsbruggen, 2009) on resistance at high risk. Clearly, there is scope here for research to clarify what should be found when and why.

At the other end of the continuum, people at low levels of risk should experience little or no self-threat on exposure to the information. Under such circumstances,
boosts to self-integrity may have unpredictable or even detrimental effects. For example, Brinol et al. (2007) showed that self-affirming before receiving a non-threatening message (not about health) reduced thoughtful message processing. We need to know more about the consequences of self-affirming among those at low risk and the processes involved.

Most studies reviewed by Harris and Epton used value affirmations, so the obtained positive effects on health cognition and motivation may be limited to or most pronounced for value affirmations. However, Jessop, Simmonds and Sparks (2009) recently examined differences between kindness, trait, and value affirmations. While all three promoted less defensiveness, only the traits condition promoted more requests for a sample of sunscreen than the control. Value- and esteem-based self-affirmations may, therefore, produce different effects (see also Schmeichel & Martens, 2005; Schmeichel & Vohs, 2009). If other-related affect (such as love and connectedness) is a key mediator of the effects of self-affirming, then there will also be differences between value-affirmations in their ability to promote open-mindedness if they vary in capacity to induce such affect. Self-affirmation theory currently treats these as different means to the same end (self-integrity bolstering or restoration), so exploring this issue may have important implications for the theory.

The benefits of group affirmations rather than self-affirmations (e.g., Derks, van Laar & Ellemers, 2009) for health also offer an interesting avenue for exploration.

With only one exception (Sherman et al., 2000, study 1) the studies reviewed by Harris and Epton (2009) tested the effects of self-affirming before rather than after threat. Whether order affects the impact of self-affirming on outcomes is an open question. For example, McQueen and Klein (2006) concluded that the order of presentation is not important to the positive effects of self-affirming, partly because of
short time intervals between delivery of the self-threat and the self-affirmation manipulation and also the immediacy of measurement of the dependent variables in most experiments. Consistent with this, in Sherman et al. (2000, study 1) self-affirmation produced equivalent positive benefits on message acceptance and intentions to those found in studies where the self-affirmation preceded the threat. However, order will clearly affect the processes involved and different mediators are probably responsible when self-affirmation prevents or repairs harm (McQueen & Klein, 2006). It may be that self-affirming after exposure to the message promotes reconsideration of the arguments (Sherman et al., 2000), affects the extent to which the individual relies on recall of thoughts generated while reading the message (Brinol et al., 2007), or the extent to which he or she employs deductive rather than inductive reasoning (Klein, Blier, & Janze, 2001). Moreover, delay between the self-affirmation and threat (whether before or after) may be critical. For instance, Critcher, Armor and Dunning (in press) have recently shown that, if a person has time to respond defensively to a threat, self-affirming is no longer able to reduce that defensive response.

This raises interesting questions about when and how self-affirmation undoes established defensive responses. Studies in the health domain have shown that self-affirming can undo defensiveness even among respondents with presumably well-established repertoires of defensive responses, such as cigarette smokers (Armitage, Harris, Napper, & Hepton, 2008; Harris, Mayle, Mabbott, & Napper, 2007). However, Sherman and Cohen (2006) have proposed that self-affirming may be less effective where chronically activated dissonance-arousing behaviours have led people to develop defensive repertoires, such as in heavier groups of smokers or drinkers. This is another good reason to expose groups of more chronic and perhaps older users
to self-affirmation manipulations. It also means that, in testing the moderating role of threat, researchers need to disentangle empirically the level of threat from chronic dissonance activation, perhaps by using experimentally manipulated rather than measured threats.

Other proposed moderators have yet to receive much if any attention in the literature on health-risk information. Sherman and colleagues have recently demonstrated that awareness of the consequences of self-affirming may be detrimental to its effectiveness. Both those who are naturally more aware of the likely effects of being self-affirmed and those who were explicitly told about them exhibited weaker effects in a series of studies (Sherman et al., 2009). Likewise, affirming within the same domain as the threat tends to be counter-productive, perhaps because it promotes confidence and feelings of impunity (Sherman & Cohen, 2006).

Self-affirming is also likely to affect some types of people more than others. Intriguingly, the most researched individual difference moderator in the broader self-affirmation literature, trait self-esteem, has received relatively little attention in the literature on health-related self-affirmation. There may be a difference in moderation by trait self-esteem as a function of whether self-affirming is spontaneous or manipulated (see, e.g., Boney-McCoy, Gibbons & Gerrard, 1999). Other individual differences associated with positive self regard and health-related information processing, such as dispositional optimism (e.g., Scheier & Carver, 1994), self-certainty (Wright, 2001), threat orientation (Thompson, Schlehofer & Bovin, 2006), regulatory focus (Higgins, 1998), preference for consistency (Cialdini, Trost, & Newson, 1995), need for closure (Kruglanski & Webster, 1996), consideration of future consequences (Strathman, Gleicher, Boninger, & Edwards, 1994) and
behavioural approach/avoidance (Carver & White, 1994), are obvious potential moderators worth exploring.

Finally, some of these moderators (e.g., awareness of the consequences of self-affirming or affirming within the same domain as the threat) may be especially important in determining the effectiveness of self-affirmation based interventions. We consider this later.

**Self-affirmation and health-behaviour change**

In reducing resistance to health-risk information self-affirmation removes an early obstacle to health behaviour change. However, while self-affirmed participants often subsequently express stronger intentions to change behaviour – and may even engage in more behaviours, such as leaflet taking or condom purchasing, consistent with these intentions – there is currently only limited evidence that they subsequently change their health behaviour (Harris & Epton, 2009).

Why is that? On the one hand, it could reflect the difficulties of trying to change behaviours, especially ones satisfying important personal or social needs. On the other hand, it could indicate problems with the motivation induced by self-affirming. These possibilities – which clearly have very different theoretical and practical implications – form the focus of the next few sections.

*Does self-affirming produce greater readiness to change?*

Researchers investigating whether self-affirmation induces greater interest in health behaviour change have borrowed their lead from the dominant paradigm in health psychology, which emphasises deliberative processes and pinpoints the development of sufficient readiness or intention to perform the new behaviour as pivotal (Conner & Norman, 2005). But how does self-affirmation affect intentions? While it is a relatively straightforward task to derive predictions from self-affirmation theory about
how self-affirmed people should respond to a potentially threatening message, as we move to testing the effects of self-affirming on more distal variables, such as intentions, the bases for predicting effects of self-affirming become less clear.

The most straightforward model in terms of its assumptions is the one that seems to be implicit in much of the research to date. This assumes that self-affirming promotes acceptance of a strong message but it is the message that promotes greater readiness to change behaviour. If so, the effects of self-affirming on intentions should be mediated by the changes that occur among self-affirmed participants on predictors of intentions following exposure to the information.

One problem with this model is that currently there is more evidence for positive effects of self-affirming on intentions than on predictors of intentions; however, measurement of the critical variables has been patchy (Harris & Epton, 2009). Ideally researchers should test the post-message effects of self-affirming on variables derived explicitly from one or more models of health behaviour. There is, after all, no shortage of relevant models from which to choose (see Conner & Norman, 2005). Indeed, one model proposes that bias in response to a threatening health message leads to both underestimation of the threat and overestimation of the efficacy of the recommended action (Das et al., 2003). Research to date has focussed considerably more on the effects of self-affirming on the former than the latter bias.

A second problem is that few researchers have tested for mediation. Where they have, however, there is some evidence to support the model. For example, among smokers in Armitage et al. (2008), message acceptance mediated the impact of self-affirmation on intentions to quit smoking and intentions, in turn, mediated the effect of acceptance on leaflet taking (significantly more self-affirmed than non-affirmed smokers took leaflets on how to quit). In van Koningsbruggen and Das (2009)
message derogation mediated the effect of self-affirming on intentions (of higher risk participants) to take an online diabetes test and intentions mediated the effects of self-affirming on the behaviour (clicking the link to the test) particularly among those at higher risk. In Epton and Harris (2008) response-efficacy mediated the effects of self-affirmation on subsequent behaviour.

There have also been failures to find mediation, but these are harder to interpret. In Epton and Harris (2008) self-efficacy did not mediate the effects of self-affirmation on subsequent behaviour, perhaps because of ceiling effects. In Jessop et al. (2009) measures of acceptance and defensiveness did not mediate the positive effect of self-affirming on taking a sample of sunscreen; however, in the above model acceptance should mediate intentions rather than behaviour. In Harris and Napper (2005) risk perceptions did not mediate the impact of self-affirmation on intentions; however, relationships between risk judgments and intentions are complex (Weinstein & Nicolich, 1993).

Because the path from acceptance to intentions uses established predictors of intentions, this model has the benefit of combining the self-affirmation literature with existing social cognition models of behaviour (e.g., the EPPM, Witte, 1992; the Theory of Planned Behavior, Ajzen, 1991) and we propose it as the default model for testing. Researchers should explicitly test this model and explore experimental as well as more traditional methods of assessing mediation (Spencer, Zanna, & Fong, 2005). Indeed, as well as the obvious benefits to theoretical integration and development, attention to models and theories also serves to clarify predictions. For example, not only does open-minded appraisal not inevitably result in message acceptance (it only does so if the information is persuasive), acceptance does not inevitably result in intentions to change; it does so when efficacy as well as threat is high (Witte, 1992).
Without such models researchers run the risk of expecting effects the models would not predict.

Of course, there may be other paths through which self-affirming influences behaviour. For example, self-affirming may induce stronger intentions to change without concomitant changes on the usual predictors, which may be why there is more evidence for main effects of self-affirming on intentions than on predictors. More generally, in addressing the “downstream” effects of self-affirmation there is a need to disentangle (both theoretically and empirically) those effects of the manipulation on predictors and outcomes mediated by the message from those that are direct effects of self-affirming. There is also evidence that self-affirming alters relationships between variables (see later), so it may change how the predictors relate to intentions. Indeed, in Harris et al. (2007) self-affirming reduced the relationship between threat and intention; it would be useful to assess whether it moderates the relationship between intention and other predictors of intentions (and the consequences of this for behaviour).

Moreover, researchers are uncovering evidence of the distinctive processes involved in the successful translation of intentions into behaviour (Schwarzer, 1992; Sheeran, Milne, Webb & Gollwitzer, 2005) and beyond into maintenance and habit (Rothman & Salovey, 2007). Attention needs to be paid to the potential impact of self-affirming at all phases, not just intention formation and action initiation.

Finally, behaviour is impulsive as well as rational. There is considerable scope for exploring whether and in what ways self-affirming affects more impulsive behaviour (e.g., Hofmann, Friese & Wiers, 2008).

So, why is there only limited evidence to date about effects on behaviour?
Of course, from the applied perspective the big question mark over self-affirmation concerns its potential for inducing health-behaviour change. Despite one recent success (Epton & Harris, 2008), most published studies testing for effects of self-affirming on health behaviour have found none, even though self-affirming has been shown to change important behaviours in non-health domains (Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009).

Epton and Harris (2008) discuss some reasons for this. The studies reporting failures have involved attempts to reduce or terminate health-compromising behaviours, such as alcohol or cigarette consumption, that require a range of physiological, lifestyle and social adjustments that complicate the process of health behaviour change; so the type of behaviour being targeted may moderate the effectiveness with which self-affirming promotes behaviour change. Health-promoting behaviours, such as improving diet or increasing exercise, may make more responsive targets (Epton & Harris, 2008). Moreover, the studies were not interventions, but experiments with brief manipulations, usually in laboratory settings, and in most the behavioural goal was vague and the message lacked information about how to achieve it. The time scale for following up the behaviour may also have been insufficient – the longest has been one month (Harris & Napper, 2005). Given this, it is perhaps less surprising that these studies failed to find effects on their targeted behaviours, despite positive changes in intentions in the self-affirmed group (e.g., Harris & Napper, 2005).

On the other hand, it may be that there are problems with the intentions formed by self-affirmed participants. Possibly they are too weak or unstable to sustain subsequent behaviour, perhaps because they are induced by heightened responsiveness to experimental demand or undue optimism. The strength and stability
of intentions are important determinants of subsequent behaviour change (Cooke & Sheeran, 2004). Indeed, if self-affirming changes how intentions are formed, this may affect such key properties of these intentions and in the process their chances of being translated into effective action. This may be one reason why there is limited evidence for effects of self-affirming on behaviour.

In fact, self-affirming could even prove counter-productive to health behaviour change: It may, for example, prompt premature attempts to change – perhaps because of ill conceived intentions, inflated perceptions of ability to enact the behaviour or barrier underestimation – that result in subsequent failure and demoralisation. Much more research attention needs to be paid to these possibilities.

However, even when self-affirming results in strong, stable and well-formed intentions, we should not expect them to translate inevitably into successful attempts at health behaviour change; common sense, as well as decades of research, tell us that there can be a gap between even the best of intentions and behaviour (Sheeran, et al., 2005). There are currently no theoretical grounds for expecting the intentions formed after self-affirming to be any better able to overcome this intention-behaviour gap than those formed otherwise.

Currently, therefore, we have no clear basis on which to establish if self-affirmation has rarely promoted health behaviour change because of the difficulties of trying to change the targeted behaviours or because of deficiencies in the motivation it induces. This is a key question to tackle in future research.

**Using self-affirmation in interventions**

Nevertheless, researchers may be encouraged to press ahead with intervention studies by the promising effects of self-affirmation on resistance and motivation. In doing so they should, of course, take note of the above issues and also the limitations of the
existing studies, which have been mainly laboratory based, used samples with high literacy and intelligence, and manipulations involving essay writing or values scales that have to be administered individually.

How to get people to self-affirm is clearer in some applied contexts – such as those involving individual counselling or small-scale group work in controlled settings (see, e.g., Charlson et al., 2007) – than in others. In particular, it is not yet clear how self-affirmation might be used in mass communications. There have been some attempts to incorporate self-affirmation manipulations into warnings and leaflets, but these have met with mixed success: the attempt by Dillard, McCaul, and Magnan (2005) was unsuccessful, but in Jessop et al. (2009) the relevant (positive traits) condition successfully promoted requests for a free sample of sunscreen.

There are also other issues that pose potential problems for interventions involving self-affirmation. McQueen and Klein (2006) point out that applications will need less convoluted cover stories, but what if the intervention signals to the recipient the purpose of self-affirming, given that such awareness appears to eliminate the benefits (Sherman et al., 2009)? Indeed, Sherman and Hartson (in press) go so far as to suggest that the “key to an effective affirmation intervention may lie in the subtlety of its delivery and the minimalism of its administration” (p. 34). Outside the laboratory it may be natural for people to focus on characteristics and attributes related conceptually to the targeted domain, yet same-domain affirmations typically backfire (Sherman & Cohen, 2006). What happens to low threat participants exposed to the message? Earlier we discussed the possibility that there may be unpredictable or even deleterious consequences for such participants, such as unwarranted boosts to judgmental confidence. There are also data (Harris & Napper, 2005) showing that self-affirming can lead to reductions in risk perceptions for threats not targeted in the
message, suggesting potentially detrimental effects on unrelated events even among high-threat participants. Self-affirmed participants also sometimes employ less effective reasoning strategies, such as when testing the validity of an hypothesis (Munro & Stansbury, 2009).

Researchers committed to intervening nonetheless, might consider whether self-affirming should be bolstered by techniques known to enhance the translation of intentions into behaviour, such as the formation of implementation intentions (Sheeran et al., 2005). It may also be useful to bolster the initial self-affirmation subsequently (see, e.g., Cohen, et al., 2009).

**The experimental paradigm**

So far the published research on the health implications of self-affirming has been exclusively experimental. While this brings many benefits, it has to date created two lacunae in the evidence base. First, it has led to a relative failure to examine whether self-affirming affects the relationships between variables (McQueen & Klein, 2006). Second, there has been little or no exploration of spontaneous rather than forced self-affirmation.

**Relationships between variables**

There is some evidence that self-affirming may change the relationships between variables. For example, in Sherman et al. (2000, study 1) the more the message evoked positive than negative thoughts, the more self-affirmed (but not non-affirmed) participants accepted it. Among self-affirmed participants in van Koningsbruggen and Das (2009), diabetes risk was positively related to intentions to take the online diabetes test and unrelated to message derogation; in contrast, in non-affirmed participants, it was unrelated to intentions and positively related to message derogation.
derogation. In Harris et al. (2007) self-affirming reduced the relationship between threat and intention.

Indeed, one implication of the idea that self-affirming uncouples the threat from the self (Sherman & Hartson, in press) is that there should be weaker correlations between measures of self-evaluation and measures related to the threatened domain in self-affirmed conditions. Sherman and Hartson describe several examples where this is the case. For example, in Sherman et al. (2009) ratings of information criticising a controversial baseball player (Barry Bonds) were negatively correlated with ratings of identification with his team (the San Francisco Giants) among non-affirmed but not among self-affirmed Giants fans. If Sherman and Hartson are right, self-affirming should similarly attenuate relationships in the health domain where ratings in the non-affirmed condition are determined primarily by defensiveness.

Klein and colleagues have also proposed that self-affirming changes the basis on which people make judgments when threatened, rendering them more deductive (i.e., less data-driven). For example, Klein et al. (2001) examined the relationship between risk judgments (for heart disease and alcohol poisoning) made two months apart in a group of high school students. On the second occasion, participants were either self-affirmed or not and either received or did not receive challenging information (in the form of the group mean for each risk factor from time 1). Klein et al. found that self-affirmed participants, when threatened, paid less attention to their behaviour in making their judgments; for example, time 1 risk factor ratings significantly and strongly predicted time 2 risk factor ratings in all groups except self-affirmed/threatened participants. More recently, Klein and Monin (2009, p. 382) report data showing that self-affirming reduced the positive relationship between
reported alcohol consumption and perceptions of breast cancer risk (in data collected but not reported by Klein & Harris, 2009).

Although both sets of ideas suggest that self-affirming should attenuate certain relationships, the protagonists view the consequences quite differently. Sherman and Hartson see this as one of the ways in which self-affirming reduces defensiveness; Klein, on the other hand, suggests that more deductive judgments can be more defensive (e.g., Klein et al., 2001). Currently, it is not clear how these differences are to be resolved. Indeed, downplaying personal relevance or susceptibility is thought to be a common defensive strategy. This should result in a weakened relationship between risk assessments and personal judgments among at-risk, non-affirmed participants, so that one effect of self-affirming should be to improve the correspondence between risk assessments and personal judgments. Yet where researchers have examined this issue, it appears the opposite is the case.

Clearly, self-affirming has the potential to change relationships between variables. More studies should examine this aspect of their data and the possibility that self-affirming changes the basis on which judgments are made. It would also be useful to know how stable any such effects are. Above all, we need to clarify theoretically where, when and why self-affirming changes relationships.

Spontaneous self-affirmation

The research reviewed so far requires participants to self-affirm. Very little empirical work has examined spontaneous self-affirmation. In the health domain there is some related research on compensatory self-enhancement (CSE), a defensive response to threatening information that involves coping by focusing on positive personal qualities; CSE has been shown to be associated with more, rather than less, defensive
responding to the health threat, at least in high self-esteem participants (e.g., Boney-McCoy et al., 1999).

There is much to discover about whether, when and how people self-affirm in their everyday lives and the consequences of doing so, both in general and in relation to health-threats. In this respect the data lag considerably behind the theoretical analyses – including that originally offered by Steele (1988). Sherman and colleagues (e.g., Sherman & Cohen, 2006) suggest that in everyday life effective self-affirmation – the sort that helps people confront threats – may operate with subtlety and without conscious intent, setting it apart from the conscious and deliberate strategies often encouraged for those low in self-regard. This is an intriguing possibility that awaits empirical scrutiny. It also remains to be seen to what extent people spontaneously self-affirm pre-emptively in the way encouraged by the manipulation, rather than reactively in the way typically described in accounts of Self-Affirmation Theory (e.g., Steele, 1988).

**In conclusion**

While we regard the picture emerging from this nascent literature as encouraging, more needs to be discovered about how self-affirming achieves its effects and their limits. We hope the material and issues covered in this review will assist researchers in developing the studies that will contribute to the next phase of the research and, in conjunction with Harris and Epton (2009), provide those who simply wish to know more about the topic with the information they need.

Health messages are powerful, often containing information about threats to life; most people are highly involved with their health and many approach health-risk information with experience in resisting elements of the message or with scepticism about health messages more generally. These features set health-risk information
apart from some of the other topics used in research on persuasion. We suggest it provides particularly fertile territory in which to investigate the effects of self-affirmation, develop its theoretical base and test its applied implications.
The impact of self-affirmation p.27

References


facilitate healthy behaviours in patients with cardiopulmonary diseases: Rationale, trial design, and methods. *Contemporary Clinical Trials, 28*, 748-962.


