Accepted Manuscript

Do Self-Incentives Change Behavior? A Systematic Review and Meta-Analysis

Emma M. Brown, Debbie M. Smith, Tracy Epton, Christopher J. Armitage

PII: S0005-7894(17)30100-4
DOI: doi:10.1016/j.beth.2017.09.004
Reference: BETH 746

To appear in: *Behavior Therapy*

Received date: 12 August 2016
Accepted date: 3 September 2017

Please cite this article as: Brown, E.M., Smith, D.M., Epton, T. & Armitage, C.J., Do Self-Incentives Change Behavior? A Systematic Review and Meta-Analysis, *Behavior Therapy* (2017), doi:10.1016/j.beth.2017.09.004

Do Self-Incentives Change Behavior? A Systematic Review and Meta-Analysis

Authors
Emma M. Brown[1]; Debbie M. Smith[2]; Tracy Epton[3]; Christopher J. Armitage[3]

Affiliations
[1] School of Health Sciences, Faculty of Medical and Health Sciences, University of East Anglia, Norwich, UK. [2] School of Social and Health Sciences, Leeds Trinity University, Leeds, UK. [3] Manchester Centre for Health Psychology, Division of Psychology and Mental Health, School of Health Sciences, Faculty of Biology, Medicine, and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK.

Corresponding Author
Professor Christopher J. Armitage, Manchester Centre for Health Psychology, Division of Psychology and Mental Health, School of Health Sciences, Faculty of Biology, Medicine, and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK.

Date of resubmission: 17/08/2017

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Abstract

Encouraging people to self-incentivize (i.e., to reward themselves in the future if they are successful in changing their behavior) or self-reward (i.e., prompt people to reward themselves once they have successfully changed their behavior) are techniques that are frequently embedded within complex behavior change interventions. However, it is not clear whether self-incentives or self-rewards *per se* are effective at bringing about behavior change. Nine databases were searched alongside manual searching of systematic reviews and online research registers. One thousand four hundred papers were retrieved, spanning a range of behaviors, though the majority of included papers were in the domain of "health psychology". Ten studies matched the inclusion criteria for self-incentive but no studies were retrieved for self-reward. The present systematic review and meta-analysis is therefore the first to evaluate the unique effect of self-incentives on behavior change. Effect sizes were retrieved from seven of the ten studies. Analysis of the seven studies produced a very small pooled effect size for self-incentives (k=7; N=1,161), which was statistically significant, d_+=0.17, CI=0.06 to 0.29. The weak effect size and dearth of studies raises the question of why self-incentivizing is such a widely employed component of behavior change interventions. The present research opens up a new field of inquiry to establish: (a) whether or not self-incentivizing and self-rewarding are effective behavior change techniques, (b) if self-incentives and self-rewards need to be deployed alongside other behavior change techniques, and (c) when and for whom self-incentives and self-rewards could support effective behavior change.

Keywords: self-incentive, self-reward, isolation, behavior change, intervention

Research on learning theory shows that incentivizing or rewarding progress towards a particular goal is likely to lead to desirable behavior change, particularly when the individual has prior knowledge that this incentive will be administered (Skinner, 1953). Positive reinforcement in the form of externally administered incentives (e.g., money, vouchers, social recognition; coded as 10.8 "incentive" in the Behavior Change Technique Taxonomy v1 [BCTT v1], Michie et al., 2013) has shown promise in changing health related behavior (Giles et al., 2014; Mantzari, Vogt, Shemilt, Wei, Higgins, & Marteau, 2015). However, the use of externally administered incentives has been associated with two main limitations, specifically: (a) the logistical difficulties associated with externally administered incentives (i.e., agreement and acceptability of these incentives to participants and the general public; Hoddinott et al., 2014); and (b) the possibility that externally administered incentives might undermine intrinsic motivation (once the incentive is removed, the likelihood of long-term behavior change is decreased and relapse may be imminent; Deci & Ryan, 1985; Gneezy, Meier, & Rey-Biel, 2011). The present research provides the first systematic review and meta-analysis to assess whether self-administered incentives ("self-incentives", BCTT v1 10.7) and self-administered rewards ("self-rewards", BCTT v1 10.9) are a viable alternative to rewards and incentives that are administered externally.

Self-incentivizing involves encouraging people to plan to incentivize themselves contingent on a pre-specified goal, whereas self-rewarding involves prompting people to reward themselves after they have achieved an important goal. Both self-incentive and self-reward overcome some of the limitations of externally administered incentives because key features of the incentive or reward (e.g., magnitude, schedule, type) are under the control of the individual. Thus, self-incentives and self-rewards are: (a) likely to be more acceptable than externally administered incentives, (b) less likely to undermine intrinsic motivation (e.g.,

Deci & Ryan, 1985), and (c) more likely to lead to desirable behavior change than externally administered incentives (Skinner, 1953).

The positive features of self-incentives and self-rewards (e.g., likely to be more acceptable) has influenced the frequent deployment of these behavior change techniques alongside several others within complex behavior change interventions across a range of behaviors including: Smoking cessation (Belles & Bradlyn, 1987), weight loss (Sciamanna et al., 2011), and self-management of pain, sleep quality and daily living activities related to juvenile fibromyalgia (Degotardi et al., 2006). Reference to, and promotion of, self-incentivizing and self-rewarding as effective behavior change techniques is common in guidance administered to practitioners charged with helping people to lose weight (Irish Nutrition and Dietetic Institution, 2013), quit smoking (Michie & Abraham, 2004) and in health behavior change more broadly (Michie et al., 2008).

The behavior change technique taxonomy v1 (BCTT v1, Michie et al., 2013) makes explicit a distinction between "self-incentives" and "self-rewards". According to the BCTT v1 (Michie et al., 2013) a "self-incentive" is defined as a "plan to reward self in future if and only if there has been effort and/or progress in performing the behaviour" (BCTT v1 10.7) whereas a "self-reward" is defined as to "prompt self-praise or self-reward if and only if there has been effort and/or progress in performing the behaviour" (BCTT v1 10.9). Thus, self-incentives are plans to reward one's self in the future if one is successful in changing one's behavior whereas self-rewards are prompts to reward oneself after having successfully changed one's behavior. However, historically, the terms "self-incentive" and "self-reward" have been used interchangeably. For example, a systematic review investigating the effectiveness of tailored programs to increase physical activity found that all 18 qualifying studies incorporated self-rewards alongside other behavior change techniques such as goal setting (BCTT v1 1.1) and building social support (BCTT v1 3.1) (Kahn et al., 2002).

However, what had been defined as self-rewards in this review, would better fit the BCTT v1 (Michie et al., 2013) definition of a self-incentive (BCTT v1 10.7).

Despite the popularity of self-incentivizing or self-rewarding as a feature of behavior change interventions, their use may be described as more "evidence-inspired" than "evidence-based" because very little is known about the unique effects of self-incentives or self-rewards on behavior change (Michie & Abraham, 2004). This is because self-incentives and/or self-rewards are typically deployed alongside several other behavior change techniques and contrasted with "treatment as usual" comparator conditions. Thus, it is not yet clear to what extent successful behavior change can be attributed to self-incentivizing or self-rewarding *per se*, which is important because self-incentivizing or self-rewarding may exert null or even negative effects and are therefore worthy of examination in their own right.

We therefore decided to undertake the first systematic review of the literature to evaluate the effectiveness of self-incentivizing and self-rewarding (independently) on behavior change. This systematic review has two key objectives, which are to calculate the effect size associated with the effects of self-incentivizing and self-rewarding (independently) and to try to understand when and for whom self-incentives or self-rewards might change behavior.

Method

This systematic review was conducted in accordance to guidelines provided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyzes (PRISMA) statement (Moher, Liberati, Tetzlaff & Altman, 2009).

Eligibility Criteria

Both published and unpublished randomized controlled trials were eligible for inclusion if they fulfilled the following criteria:

1. Consisted of human participant populations with no limitations for age, medical conditions or language the manuscript was published in.
2. The self-incentivizing or self-rewarding condition could be compared to an alternative intervention (e.g., self-monitoring, BCTT v1 2.3), if self-monitoring was also used in the intervening condition), usual care, or no other intervention.
3. Self-incentives or self-rewards were utilized in isolation of additional behavior change techniques, or could be controlled as the only difference between the intervention and comparator conditions[1].
4. The self-incentive or self-reward and kinds of incentives or rewards (e.g., material incentive, BCTT v1 10.1) chosen are contained only within the intervention condition and are not found in the comparator condition.
5. The outcome measure was behavior with no limitations in terms of the target behavior.
6. Behavior was assessed using self-report and/or objective measures at least one point in time following (including immediately following) the intervention/comparator.

Data Sources and Search Strategy

Electronic databases were searched without date limits up to the date of the search (December, 2015). These included: The Allied and Complementary Medicine Database (AMED), Applied Social Sciences Index and Abstracts (ASSIA), Cumulative Index to Nursing and Allied Health Literature plus (CINAHL), Cochrane Central Register of Controlled Trials (CENTRAL), Embase, Health and Psychosocial Instruments (HaPI), MEDLINE, PsycINFO, Science Citation Index Expanded (SCI-EXPANDED), and Social Sciences Citation Index (SSCI). Several databases were searched due to a dearth of studies matching the inclusion criteria.

The search strategy utilized relevant search terms for an incentivizing or rewarding aspect (e.g., self-treat, self-incentive, self-reward [BCTT v1 10.9]; see supplementary material S2-S5 for the search strategies used). Manual searching of relevant systematic reviews, online research registers (clinicaltrials.gov, ISRCTN), and general web searching was undertaken to identify further studies that matched the inclusion criteria. The reference lists and citation searches of all the included studies were also screened for additional relevant studies.

Study Selection and Data Extraction

After the removal of duplicates, one researcher (EMB) independently screened the titles and abstracts of all retrieved papers to exclude those that were irrelevant to the eligibility criteria. Full texts of the remaining papers were obtained and independently screened by one researcher (EMB) to identify those that met the inclusion criteria (see supplementary material S7 for a list of excluded papers with reasons). Additional researchers (TE & DMS) independently screened 10% of the titles and abstracts, achieving very good agreement (Cohen's Kappa >0.8; Peat, 2001) and 10% of the full texts, resulting in moderate agreement (Cohen's Kappa = 0.5). This lower Cohen's Kappa value at full-text stage is representative of the low number of studies screened at this stage (26 studies) with only one disagreement. Any disagreements at title, abstract, and full-text stage were resolved through discussion, and a third reviewer was not required.

Data were extracted from all included studies by one researcher (EMB) using a data extraction sheet created for this purpose. Additional researchers (TE & DMS) also extracted data from all included studies for accuracy, resulting in above moderate agreement (Cohen's Kappa = 0.7) with any disagreements that arose being resolved through discussion. Information extracted from the included studies consisted of: Sample characteristics, study design, interventions used, all outcomes, and results. Data were also extracted if available to

address questions about the effectiveness of self-incentivizing or self-rewarding in relation to compliance, administration, and duration.

The nature of self-incentivizing and self-rewarding was coded using Adams, Giles, McColl, and Sniehotta's (2014) framework for describing financial incentive interventions. They consisted of: (1) how frequently the instances of behavior were rewarded/incentivized, (2) how immediately the reward/incentive was received, and (3) if the reward/incentive was the same throughout or if this varied (e.g., increased in value as the behavior was maintained).

Risk of Bias

Risk of bias was rated for all included studies by one researcher (EMB) following criteria recommended by the Centre for Reviews and Dissemination, (2009) (see supplementary material S8). A second researcher (TE) rated all of the included studies to check for agreement, which resulted in moderate agreement (Cohen's Kappa = 0.6). Any disagreements at this stage were resolved through discussion.

Data Synthesis

The included studies were grouped together in order of relevant psychological field (based on the divisions of the American Psychological Association) in a tabular form for narrative synthesis (see supplementary material S7). If studies reported more than one relevant comparison condition, then the comparison based on no intervention was favored, with the other comparisons being excluded as irrelevant. For example, the included paper by Castro and Rachlin (1980) had four possible conditions, which included: (1) self-incentive (BCTT v1 10.7), (2) self-monitoring (BCTT v1 2.3), (3) self-punishment, and (4) control, in which the self-incentivizing and the control conditions were analyzed for the present review.

If it was not possible to compare the intervention condition with a no intervention control, then a comparison condition such as the utilization of another behavior change technique was included. For example, the paper by Chapman and Jeffrey (1978) had three

possible conditions, which included: (1) self-monitoring (BCTT v1 2.3), (2) self-monitoring plus goal setting (BCTT v1 2.3 plus 1.1), and (3) self-monitoring plus goal setting plus self-incentive (BCTT v1 2.3 plus 1.1 plus 10.7). In this instance, conditions two (self-monitoring plus goal setting), and three (self-monitoring plus goal setting plus self-incentive) were analyzed for the present review, as the only difference between these conditions was the inclusion of a self-incentive. Analysis of Chapman and Jeffrey's (1978) and related studies in this way ensured that the effect size could not be attributed erroneously to a behavior change technique other than self-incentive.

Differences between the delivery (e.g., immediately self-incentivizing/rewarding, self-incentivizing/rewarding weekly) and choosing of the self-incentive or self-reward (e.g., instructed how to self-incentivize/reward) were also included where available. Where multiple follow-up time points were reported, only the post-intervention time point was used to calculate the effect sizes to be pooled within the meta-analysis to allow for the most appropriate representative effects based on the self-incentivizing or self-rewarding interventions. All but two studies did not report an additional follow-up time point once the intervention had ceased, or had used a delayed intervention for the control group at which point the control group was no longer an appropriate comparator.

Meta-Analytic Strategy

The following procedures were implemented to compute the standardized mean difference (i.e., Cohen's *d*, Cohen, 1992) for each study included in this meta-analysis in order to calculate the overall pooled effect size. A bias corrected function (Hedges & Olkin, 1985) was used across all included studies to correct for bias in standardized mean differences, which has been previously suggested for studies that contained samples of fewer than 20 participants (Hedges & Olkin, 1985).

For each study comparison, effect sizes were calculated based on the available data (see supplementary material, S6). The preferred method of effect size calculation was to subtract

the post-test mean from the baseline mean for each intervention and comparator condition, and divide the result by the pooled baseline standard deviations of the two conditions. The statistics required for the preferred method of effect size calculation (namely, baseline and post-test means, standard deviations, and sample sizes for all conditions) were available for three of the ten included studies (Armitage, 2014; French et al., 1994, two separate analyses).

When these data were not available, the second preferred method of effect size calculation was to use tables that compared the frequencies of participants whose behavior did change with those whose behavior did not change for both the intervention and comparator conditions at post-treatment. This method uses the frequencies as outlined above for both the intervention and comparator conditions to calculate an odds ratio. From this, the odds ratio is converted in to Cohen's *d* by multiplying the natural logarithm of the odds ratio by the square root of 3, and dividing this value by the number Pi ($d = \ln OR \sqrt{3}/\pi$; Hasselblad & Hedges, 1995). The statistics required for this method were available in two of the ten included studies (Hailey et al., 1992; Solomon et al., 1998).

The final and least preferred method of effect size calculation used the one-tailed *p* value reported for the intervention and comparator condition alongside the sample size. For this calculation, the z value associated with the one-tailed *p* value was found (using a *z*-table provided by Field, 2009). From this, using the method proposed by Rosenthal (1991), the *z* value was divided by the square root of the total number of participants across both the intervention and comparator conditions to calculate Pearson's *r*. Finally, the method proposed by Friedman (1968) of two multiplied by *r*, divided by the square root of the term $(1 - r^2)$ was used to convert Pearson's *r* in to Cohen's *d* ($d=2r/\sqrt{1-r^2}$). Two of the ten included studies provided only the one-tailed *p* values to calculate the effect size between the intervention and comparator conditions to allow for the investigation of the effectiveness of self-incentivizing in isolation (Castro & Rachlin, 1980; Jackson & Molloy, 1985). The bias corrected function

(1 - 3/(4(intervention sample size+control sample size-2)-1)) as proposed by Hedges & Olkin, 1985) was used for all of the above calculations to correct for any sample size biases present.

The remaining three studies lacked the reporting of any data (e.g., means, standard deviations, frequencies, *F* values, *t* values, or *p* values) that could be used for the accurate calculation of effect size between the intervention and comparator conditions (Chapman & Jeffrey, 1978; Grady, Goodenow, & Borkin, 1988; Greiner & Karoly, 1976). All authors were contacted and asked to provide access to these additional data, but the data could not be retrieved. Therefore, the remaining three studies were analyzed in the following three ways: (a) omitting these studies, (b) assuming that the effect size was zero, and (c) calculating an effect size assuming that $p = .5$.

A random-effects model, weighted by sample size was used to calculate the pooled effect size, confidence intervals, significance of heterogeneity, and extent of heterogeneity for all outcome variables, using the revised metan command in STATA Version 12.1 (Stata, 2011). Random-effects models were conducted due to differences in the included studies (e.g., effect sizes) which were expected to be heterogeneous. Levels of heterogeneity were assessed through the I^2 statistic based on suggestions that an I^2 over 50% equates moderate heterogeneity, and over 75% equates high heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003).

Potential moderators of self-incentivizing or self-rewarding (e.g., type of self-incentives/rewards chosen; schedule of delivery) could not be explored due to the dearth of studies that were eligible for inclusion in the present systematic review.

Results

Study Characteristics

One thousand four hundred papers were retrieved. After the removal of duplicate papers, the titles and abstracts of the remaining 1255 papers were screened: 990 papers were excluded at this stage, as they did not meet the inclusion criteria. Full-texts of the remaining

265 papers were screened and nine papers that met the inclusion criteria were included in the present review (see Figure 1 for PRISMA flow diagram, including reasons for exclusion). One of the included papers (French et al., 1994) incorporated two self-incentivizing intervention conditions and two comparator conditions with the only difference between the intervention and comparator conditions being self-incentive. Taking these additional conditions in to account within the study, overall ten self-incentivizing comparisons were analyzed.

Although we included both "self-incentive" and "self-reward" as search terms in the present systematic review, none of the interventions described in the retrieved comparisons could be accurately described as "self-reward" (i.e., people are prompted to self-reward after the behavior has already changed) but did meet the definition of a "self-incentive" (i.e., people are asked to plan to reward themselves in future if they are successful in changing their current behavior) according to the BCTT v1 (Michie et al., 2013).

Risk of selection bias was low across all included studies. There was no allocation bias in one study (Armitage, 2014), which reported all randomization procedures and concealed condition allocation. Allocation bias was high for the remaining nine studies. Detection bias ranged from moderate to high across all included studies, due to a lack of information about whether the assessors, and/or those who administered the intervention were blind to condition and also how blinding was assessed. Attrition bias was low for one study (Armitage, 2014), with the remaining studies lacking information about participant withdrawals, and the use of intention to treat analysis.

The sample sizes for four out of the seven studies meta-analyzed within the present systematic review were considered small (ranged between 20-32 participants per condition) and for this reason, the meta-analysis was weighted by sample size. Social desirability bias may also be present for six of the seven studies analyzed because self-report measures were

used to assess behavior change. Research in behavior change has been heavily reliant on self-reported outcomes (Cascio, Dal Cin & Falk, 2013; Webb & Sheeran, 2006) due to underlying issues of cost-effectiveness, feasibility and ethical considerations of objectively assessed outcomes (Brener, Billy, & Grady, 2005). Whilst subjectively assessed outcomes are not ideal due to the potential to provide inaccurate information, self-reporting of some behaviors (e.g., smoking) has been previously suggested to be highly accurate (Velicer, Prochaska, Rossi & Snow, 1992).

Narrative Synthesis

Of the ten included studies, three studies focused on breast self-examination (Grady et al., 1998; Hailey et al., 1992; Solomon et al., 1988), two studies focused on weight loss/control (Castro & Rachlin, 1980; Chapman & Jeffrey, 1978), two on physical activity (French et al., 1994, two separate analyses), one on fruit consumption (Armitage, 2014), one on academic performance (Greiner & Karoly, 1976), and one on task performance of arithmetic problems (Jackson & Molloy, 1985). Four studies recorded where participants had been recruited from, with three studies that collected data from the USA (Grady et al, 1998; Hailey et al, 1992; Solomon et al, 1988), and one from Romania (Armitage, 2014). The remaining data were likely collected in the authors' countries of origin, namely, the USA (Castro & Rachlin, 1980; Chapman & Jeffrey, 1978; French et al., 1994, two separate analyses; Greiner & Karoly, 1976), and Australia (Jackson & Molloy, 1985).

If the type of self-incentives chosen by the participants was provided within the included papers, then these were also coded as material (BCTT v1 10.1), social (BCTT v1 10.5), or non-specific (BCTT v1 10.6, Michie et al., 2013). The types of self-incentives that participants planned to implement were unclear in seven of the ten studies because participants were asked to self-generate their own self-incentive. Two studies asked participants to use money as the (material) self-incentive (Castro & Rachlin, 1980; Jackson &

Molloy, 1985) and one study (Armitage, 2014) asked participants to self-generate self-incentives that were either material (e.g., shopping) or social (e.g., meeting friends). Authors of the remaining seven studies were contacted in an attempt to provide sufficient data about the type of self-incentive that participants had chosen. However, this information was no longer available due to the prolonged period of time since the data were originally collected.

The schedule of self-incentivizing varied across the ten studies. One study asked participants to self-incentivize immediately after one instance of the target behavior (Jackson & Molloy, 1985), five studies asked participants to plan weekly self-incentives (Armitage, 2014; Castro & Rachlin, 1980; French et al., 1994, two separate analyses; Greiner & Karoly, 1976) contingent on successful changes in behavior. One study asked participants to plan weekly self-incentives for changes in eating behavior and monthly self-incentives for changes in physical activity (Chapman & Jeffrey, 1978). The final three studies asked participants to self-incentivize on a monthly basis (Grady et al., 1988; Hailey et al., 1992; Solomon et al., 1998) immediately after breast self-examination.

The use of self-incentives by participants was not specified in four studies (Castro & Rachlin, 1980; Chapman & Jeffrey, 1978; Greiner & Karoly, 1976; Jackson & Molloy, 1985). Three studies mentioned that participants used self-incentives equally across all conditions, although participants within the comparator conditions had not been encouraged to self-incentivize (Armitage, 2014; French et al., 1994, two separate analyses). The remaining three studies recorded participants' use of self-incentives within the intervention conditions at 11% (Solomon et al., 1998), 39% (Hailey et al., 1992), and 51% (Grady et al., 1988) with only one study reporting the spontaneous administration of self-incentivizing in the comparator condition (5%, Solomon et al., 1998). The final and only follow-ups for the majority of the included studies ranged from immediately following a problem-solving task in one study (Jackson & Molloy, 1985) to 12-months post intervention (Solomon et al.,

1998), *M* = 20 weeks. Whilst three studies reported a further follow-up after the initial follow-up, only two studies (French et al., 1994, two separate analyses) measured and reported additional further follow-up outcomes of which relevant comparisons could be analyzed. The additional study (Castro & Rachlin, 1980) that reported a further follow-up used a delayed comparator condition that incorporated self-incentives in both the intervention and comparator conditions after the initial follow-up measures had been recorded and thus would no longer contain a relevant comparator to analyse in the present meta-analysis. The lack of additional follow-up measures within the included studies determined the use of only initial follow-up or post-intervention measures to calculate the effect sizes within the present systematic review and meta-analysis.

Of the ten included studies, five studies produced statistically significant results for the self-incentivizing intervention condition compared to a relevant comparator. Those that produced significant results (Armitage, 2014; Castro & Rachlin, 1980; French et al. 1994; Jackson & Molloy) utilized self-incentivizing based interventions across a range of populations (students, general population), target behaviors (increase fruit consumption, weight loss, physical activity, and problem-solving), and alongside additional behavior change techniques (e.g., self-monitoring) utilized within both intervention and comparator conditions.

Overall effect of self-incentive

The pooled effect size for self-incentives (see Table 1 for individual effect sizes), excluding those studies from which effect sizes could not be extracted (k = 7) and including 1,161 participants overall, suggests that the effect of self-incentivizing is very small (between 0.1 and 0.2, Sawilowski, 2009) and statistically significant (d_+ = 0.17, *CI* = 0.06 to 0.29, I^2 = 0.0%; Q = 4.05); see S1 of the supplementary material). The pattern of findings was similar after including three studies from which effect sizes could not be extracted and assumptions

of $d = 0$ ($k = 10$; $N = 1{,}286$, $d_+ = 0.16$, $CI = 0.05$ to 0.27; $I^2 = 0.0\%$; $Q = 4.74$), and $p = .5$ ($k = 10$; $N = 1{,}286$, $d_+ = 0.16$, $CI = 0.05$ to 0.27; $I^2 = 0.0\%$; $Q = 4.87$) were followed. Potential outliers were explored using Tukey's method boxplot 1.5 interquartile range (Hubert & Vandervieren, 2008), which is resilient to extreme values and more appropriate for small datasets. Regardless of the variations in effect sizes across the included studies, no studies emerged as outliers within this analysis.

Discussion

Self-incentives are used frequently within complex behavior change interventions but have rarely been tested without potentially confounding variables. The present research is the first systematic review to evaluate the unique effect of self-incentivizing on behavior change. The principal findings were that: (a) self-incentives produce very small effects on behavior change when used in isolation, (b) people frequently fail to administer self-incentives even when asked to do so, and (c) despite the fact that "self-rewards" are frequently referenced in the literature, it is clear that self-incentives are typically mislabelled as "self-rewards", which illustrates the importance of adopting taxonomies to describe behavior change techniques accurately (Michie et al., 2013). The following discussion considers the theoretical and clinical impact of the present findings.

The overall effect size for self-incentives was positive and statistically significant ($d_+ = 0.17$, $CI = 0.06$ to 0.29) but very small (Sawilowski, 2009). However, in comparison with previous meta-analyses that have assessed the unique effects of behavior change techniques, the number of included studies and size of effect for self-incentivizing was very small. For example, Mento, Steel, and Karren (1987) found a small mean effect size ($d_+ = 0.42$) across 49 studies utilizing goal setting (BCTT v1 1.3); Kluger and DeNisi (1996) found a small mean effect size ($d_+ = 0.41$) across 131 studies that utilized feedback on behavior (BCTT v1 2.2); Harkin et al. (2015) found a small mean effect size ($d_+ = 0.40$) across 138 studies that utilized self-monitoring (BCTT v1 2.3); and Mantzari et al. (2015) found a small

mean effect size (d = 0.34) across 34 studies utilizing externally administered incentives (BCTT v1 10.8) towards changes in habitual health related behaviors. Several alternative behavior change techniques therefore show more promise than self-incentives in terms of designing future complex behavior change interventions, however the lack of available studies and the appropriate measures within those studies (i.e., frequency of self-incentivizing) means that it was not possible to proceed with moderator analyses to evaluate or infer the effectiveness of self-incentives alongside other behavior change techniques.

Nevertheless, with just the seven self-incentivizing studies available to analyze and the popularity of self-incentivizing as a component of behavior change interventions (e.g., Kahn et al., 2002), the question arises as to why the effects of self-incentivizing are so small. It is notable that six of the ten retrieved studies made reference to participants' failure to self-incentivize as a possible implication of the effectiveness of the self-incentivizing intervention, however only one study (Grady et al., 1988) accurately measured and reported self-incentive fidelity towards the behavior change outcome.

Grady and colleagues (1988) found that rates of breast self-examination increased from 25% to 71% when self-incentives were successfully administered, with 51% of participants administering self-incentives after instruction. Unfortunately, data were unavailable to establish the effects of self-incentivizing for just those participants who did successfully administer self-incentives, and the effect size calculated for this study was based on assumptions of d = 0 and p = .5 through a lack of sufficient data between only the self-incentivizing intervention and relevant comparator condition. Additionally, Grady et al.'s (1988) study produced the highest rates of successful self-incentive administration at 51%, through monthly postal prompts, compared to an additional study that provided no means of encouragement to self-incentivize with associated rates of 11% (Solomon et al., 1998) and a very small effect size of d_+ = 0.13 (Sawilowsky, 2009).

However, Hailey and colleagues (1992) instructed participants to choose self-incentives that would be realistic for: (a) the participants to administer, and (b) for the target behavior (breast self-examination). Additionally, the participants were instructed to choose self-incentives in relation to the time and place of intended behavior performance, and self-incentive administration. Hailey et al.'s (1992) method produced a small effect size of d_+ = 0.40 with only 39% of participants complying with the self-incentive instructions (i.e., intervention fidelity). Hailey et al.'s (1992) findings suggest that poor treatment fidelity, rather than the effectiveness of self-incentivizing *per se*, might be investigated further.

Although the present review evaluates the available evidence with respect to the effectiveness of self-incentivizing to produce changes in behavior, the potential limitations should also be acknowledged. First, despite the use of very broad search terms and an extensive search strategy the low number of self-incentivizing studies retrieved coupled with the poor reporting of procedures and data reduced considerably the power of the meta-analysis. Additionally, the lack of self-rewarding interventions found within the present systematic review (as defined by the BCTT v1, Michie et al., 2013) precluded any analyses to explore the unique effect of this behavior change technique, on behavior change. This discovery provides a clear need to conduct primary research using randomised controlled trial methodology into the unique effect of self-rewarding in future research.

Second, allocation bias was high due to the lack of information on randomization and allocation, attrition bias was also moderate due to a lack of information about withdrawals and drop-outs, and detection bias ranged from moderate to high due to the poor reporting of blinding procedures. Due to the low number of included studies, excluding these studies in sensitivity analyses was not warranted. In future, better reporting procedures should be followed with a particular emphasis on issues relating to bias in allocation, attrition and detection within behavior change interventions.

Third, a further limitation is the lack of included papers from major fields of behavior modification outside of those in health psychology (e.g., clinical and consulting psychology), which unintentionally limits the generalizability of these findings. One way to remedy this is that less restrictive inclusion criteria could be followed, allowing for the inclusion of other research designs outside of solely group design randomized controlled trials within systematic reviews and meta-analyses. However, the two well-controlled single case design studies (Belles & Bradlyn, 1987; Cipani & Wolter, 1983) retrieved during the screening process were not included as it was impossible to untangle the unique effects of self-incentives on behavior change. Consequently, there appears to be a pressing need for randomized controlled trials in behavioral fields outside of health psychology including but not restricted to: clinical and consulting psychology, intellectual and developmental difficulties, organizational psychology, and social psychology (Table 2) in order to most accurately assess the effectiveness of behavior change interventions and expand the representativeness of such systematic reviews and meta-analyses.

In conclusion, the present review raises the question of why self-incentivizing is such a widely-employed component of complex behavior change interventions when there are reasons to doubt its effectiveness. The answer is that there is a dearth of primary research into the effectiveness of self-incentivizing, and indeed self-rewarding, that needs to be urgently addressed. There is considerable need for future research to establish whether: (a) the effectiveness of self-incentivizing and self-rewarding can be improved, (b) self-incentives and self-rewards need to be deployed alongside other behavior change techniques, and (c) self-incentives and self-rewards are effective for diverse populations and behaviors.

References

Adams, J., Giles, E. L., McColl, E., & Sniehotta, F. F. (2014). Carrots, sticks and health behaviors: A framework for documenting the complexity of financial incentive interventions to change health behaviors. *Health Psychology Review, 8*, 286-295. doi: 10.1080/17437199.2013.848410.

*Armitage, C. J. (2014). Evidence that self-incentives increase fruit consumption: A randomized exploratory trial among high-risk Romanian adolescents. *Prevention Science*, *15*, 186-193. doi: 10.1007/s11121-012-0346-6.

Belles, D., & Bradlyn, A. S. (1987). The use of the changing criterion design in achieving controlled smoking in a heavy smoker: A controlled case study. *Journal of Behavior Therapy and Experimental Psychiatry*, *18*, 77-82. doi: 10.1016/0005-7916(87)90075-9.

Brener, N. D., Billy, J. O., & Grady, W. R. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. *Journal of Adolescent Health*, *33*, 436-457. doi: 10.1016/S1054-139X(03)00052-1.

Cascio, C. N., Dal Cin, S., & Falk, E. B. (2013). Health communications: Predicting behavior change from the brain. In Hall, P. A. (Eds.), *Social Neuroscience and Public Health* (pp. 57-71). New York. Springer.

*Castro, L., & Rachlin, H. (1980). Self-reward, self-monitoring, and self-punishment as feedback in weight control. *Behavior Therapy*, *11*, 38-48. doi: 10.1016/S0005-7894(80)80034-7.

Centre for reviews, & dissemination (CRD). (2009). *Systematic reviews: CRD's guidance for undertaking reviews in health care*. Centre for Reviews and Dissemination. https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf. Accessibility verified July 01, 2016.

Chapman, S. L., & Jeffrey, D. B. (1978). Situational management, standard setting, and self-reward in a behavior modification weight loss program. *Journal of Consulting and Clinical Psychology, 46*, 1588-1589. doi: 10.1037/0022-006X.46.6.1588.

Cipani, E., & Wolter, J. (1983). The effectiveness of immobilization in treating self-injurious behavior. *Behavioral Engineering, 8,* 154-158.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159. doi: 10.1037/0033-2909.112.1.155.

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

Degotardi, P. J., Klass, E. S., Rosenberg, B. S., Fox, D. G., Gallelli, K. A., & Gottlieb, B. S. (2006). Development and evaluation of a cognitive-behavioral intervention for juvenile fibromyalgia. *Journal of Pediatric Psychology, 31*, 714-723. doi: 10.1093/jpepsy/jsj064.

Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.

Friedman, H. (1968). Magnitude of experimental effect and a table for its rapid estimation. *Psychological Bulletin*, 70, 245-251. doi: 10.1037/h0026258.

*French, S. A., Jeffery, R. W., & Oliphant, J. A. (1994). Facility access and self-reward as methods to promote physical activity among healthy sedentary adults. *American Journal of Health Promotion, 8*, 257-262. doi: doi.org/10.4278/0890-1171-8.4.257.

Giles, E. L., Robalino, S., McColl, E., Sniehotta, F. F., & Adams, J. (2014). The effectiveness of financial incentives for health behavior change: Systematic review and meta-analysis. *PLoS ONE, 9,* 1-16. doi: 10.1371/journal.pone.0090347.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives, 25,* 191-209. doi: 10.1257/jep. 25. 4.19.

Grady, K. E., Goodenow, C., & Borkin, J. R. (1988). The effect of reward on compliance with breast self-examination. *Journal of Behavioral Medicine, 11,* 43-57. doi: 10.1007/BF00846168.

Greiner, J. M., & Karoly, P. (1976). Effects of self-control training on study activity and academic performance: An analysis of self-monitoring, self-reward, and systematic-planning components. *Journal of Counseling Psychology, 23,* 495-502. doi: 10.1037/0022-0167.23.6.495.

*Hailey, B. J., Lalor, K. M., Byrne, H. A., & Starling, L. M. (1992). The effects of self-reinforcement and peer-reinforcement on the practice of breast self-examination. *Health Education Research, 7,* 165-174. doi: 10.1093/her/7.2.165.

Harkin, B., Webb, T. L., Chang, B. P., Prestwich, A., Conner, M., Kellar, I., ... & Sheeran, P. (2016). Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. *Psychological Bulletin, 142,* 198-229. doi: 10.1037/bul0000025.

Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin, 117,* 167-178. doi: 10.1037/0033-2909.117.1.167.

Hedges, L. V., & Olkin, I. (1985). *Statistic methods for meta-analysis*. Florida: Academic Press.

Hoddinott, P., Morgan, H., MacLennan, G., Sewel, K., Thomson, G., Bauld, L., . . . Campbell, M. K. (2014). Public acceptability of financial incentives for smoking cessation in pregnancy and breast feeding: A survey of the British public. *British Medical Journal Open, 4,* 1-9. doi: 10.1136/bmjopen-2014- 005524.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyzes. *British Medical Journal, 327,* 557-560. doi: 10.1136/bmj.327.7414.557.

Hubert, M., & Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics & Data Analysis, 52,* 5186-5201. doi: 10.1016/j.csda.2007.11.008.

Irish Nutrition and Dietetic Institution. (2013). *Steps towards a healthier weight: Behaviour Change*. Retrieved from https://www.indi.ie/fact-sheets/healthy-eating,-healthy-weight-and-dieting/434-steps-towards-a-healthier-weight-behaviour-change.html**.** Accessibility verified July 01, 2016.

*Jackson, H. J., & Molloy, G. N. (1985). Some effects of feedback alone and four types of self-consequation on selected measures of problem solving. *Perceptual and Motor Skills, 61,* 1005-1006. doi: 10.2466/pms.1985.61.3.1005.

Kahn, E. B., Ramsey, L. T., Brownson, R. C., Heath, G. W., Howze, E. H., Powell, K. E., ... & Corso, P. (2002). The effectiveness of interventions to increase physical activity: A systematic review. *American Journal of Preventive Medicine, 22,* 73-107. doi: 10.1016/S0749-3797(02)00434-8.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119,* 254-284. doi: 10.1037/0033-2909.119.2.254.

Mantzari, E., Vogt, F., Shemilt, I., Wei, Y., Higgins, J. P., & Marteau, T. M. (2015). Personal financial incentives for changing habitual health-related behaviors: A systematic review and meta-analysis. *Preventive Medicine, 75,* 75-85. doi: 10.1016/j.ypmed.2015.03.001.

Mento, A. J., Steel, R. P., & Karren, R. J. (1987). A meta-analytic study of the effects of goal setting on task performance: 1966–1984. *Organizational Behavior and Human Decision Processes, 39,* 52-83. doi: 10.1016/0749-5978(87)90045-8.

Michie, S., & Abraham, C. (2004). Interventions to change health behaviors: Evidence-based or evidence-inspired? *Psychology & Health, 19,* 29-49. doi: 10.1080/0887044031000141199.

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis J., Hardeman, . . . Wood, C. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine, 46,* 81-95. doi: 10.1007/s12160-013-9486-6.

Michie, S., Rumsey, N., Fussell, A., Hardeman, W., Johnston, M. , Newman, S. and Yardley, L. (2008). *Improving health: changing behaviour. NHS health trainer handbook.* Retrieved from http://eprints.uwe.ac.uk/12057/1/dh_085778.pdf. Accessibility verified July 01, 2016.

National Institutes of Health. (2013). *Changing Your Habits: Steps to Better Health.* Retrieved from http://www.niddk.nih.gov/health-information/health-topics/diet/changing-habits/Documents/Changing_Your_Habits.pdf. Accessibility verified July 01, 2016.

Peat, J. (2001). *Health science research: A handbook of quantitative methods*. Sydney: Sage.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* California: Sage.

Sciamanna, C. N., Kiernan, M., Rolls, B. J., Boan, J., Stuckey, H., Kephart, D., ... & Hwang, K. O. (2011). Practices associated with weight loss versus weight-loss maintenance: Results of a national survey. *American Journal of Preventive Medicine 41*, 159-166. doi: 10.1016/j.amepre.2011.04.009.

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 8, 597 – 599.

Skinner, B. F. (1953). *Science and Human Behavior*. New York: Macmillan.

*Solomon, L. J., Flynn, B. S., Worden, J. K., Mickey, R. M., Skelly, J. M., Geller, B. M., . . . Webster, J. A. (1998). Assessment of self-reward strategies for maintenance of breast self-examination. *Journal of Behavioral Medicine, 21*, 83-102. doi: 10.1023/A:1018767505776.

Stata. (2011). *Stata 12.1*. Texas: Stata Press.

Velicer, W. F., Prochaska, J. O., Rossi, J. S., & Snow, M. G. (1992). Assessing outcome in smoking cessation studies. *Psychological Bulletin, 111*, 23-41. doi: 10.1037/0033-2909.111.1.23.

Footnotes

[1] Unexpectedly, very few studies met our inclusion criteria, despite the fact that we included both "self-incentive" and "self-reward" as search terms. Although self-incentives have been frequently deployed within complex behavior change interventions, many of these studies have not followed a randomized controlled trial methodology (see Figure 1). We considered relaxing our inclusion criteria and re-ran the search to include any studies that utilized self-incentives or self-rewards alongside other behavior change techniques (i.e., self-incentives or self-rewards were not used in isolation). However, this search yielded just four further studies, none of which included the appropriate measures, such as the frequency of self-incentivizing that would allow us to proceed with moderator analysis in order to evaluate the effectiveness of self-incentives on changes in behavior. We therefore decided to retain the original inclusion criteria.

Table 1.

Effect Sizes for Self-Incentive (Studies included in the Meta-Analysis)

Study	*N* (Intervention)	*N* (Control)	*d (at post-intervention)*	*d (after intervention)*
Armitage (2014)	77	79	0.21	Not reported
Castro et al. (1980)	10	10	0.76	Delayed follow-up (no longer an accurate comparator)
French et al. (1994) (1)	14	14	-0.09	0.03
French et al. (1994) (2)	14	14	0.14	-1.19
Hailey et al. (1992)	62	54	0.40	Not reported
Jackson et al. (1985)	14	14	0.21	Not reported
Solomon et al. (1998)	373	412	0.13	Qualitative data only

Table 2.

Number of Full Text Papers both Excluded and Included within the Respective Psychological Field

Respective Psychological Field	Excluded Papers (*N*)	Included Papers (*N*)
Behavior Analysis	10	0
Clinical and Consulting Psychology	18	0
Consumer Psychology	4	0
Developmental and Educational Psychology	64	1
Experimental Psychology	31	1
Health Psychology	44	7
Intellectual and Developmental Disabilities	12	0
Organizational Psychology	41	0
Personality and Social Psychology	32	0

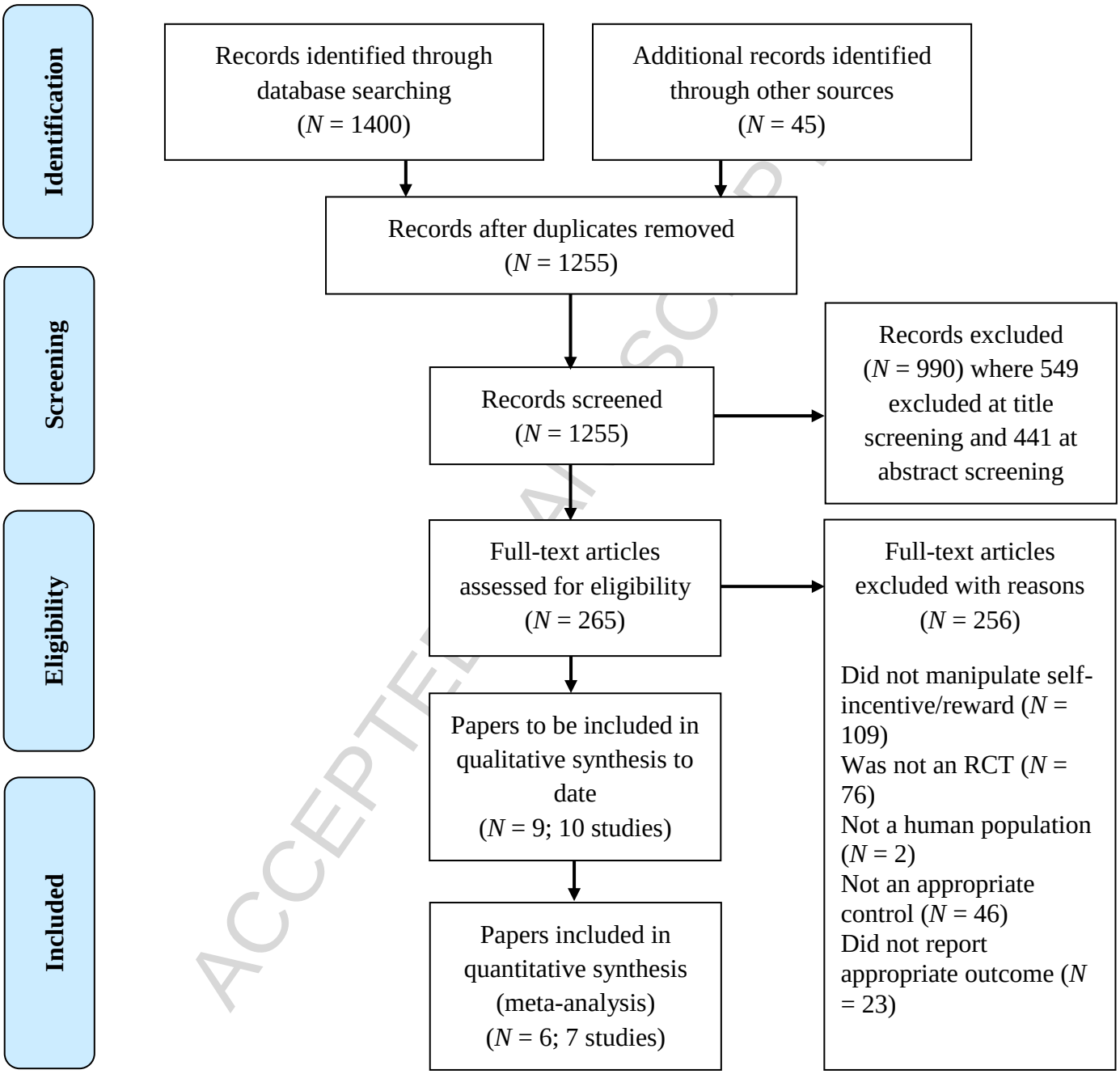

Figure 1. PRISMA flow diagram of selection and exclusion (Moher, Liberati, Tetzlaff & Altman, 2009)