Perspectives on Behavior Science (2022) 45:819–861
https://doi.org/10.1007/s40614-022-00358-7

A | B | A | I
Association for Behavior Analysis International

REVIEW

A Systematic Review of Treatment Maintenance Strategies in Token Economies: Implications for Contingency Management

Sean D. Regnier[1] · Haily K. Traxler[1] · Amanda Devoto[2] · Anthony DeFulio[3]

Accepted: 15 August 2022 / Published online: 1 September 2022

Abstract

Contingency management (CM) interventions are based on operant principles and are effective in promoting health behaviors. Despite their success, a common criticism of CM is that its effects to not persist after the intervention is withdrawn. Many CM studies evaluate posttreatment effects, but few investigate procedures for promoting maintenance. Token economy interventions and CM interventions are procedurally and conceptually similar. The token economy literature includes many studies in which procedures for promoting postintervention maintenance are evaluated. A systematic literature review was conducted to synthesize the literature on treatment maintenance in token economies. Search procedures yielded 697 articles, and application of inclusion/exclusion criteria resulted in 37 articles for review. The most successful strategy is to combine procedures. In most cases, thinning or fading was combined with programmed transfer of control via social reinforcement or self-management. Social reinforcement and self-monitoring procedures appear to be especially important, and were included in 70% of studies involving combined approaches. Thus, our primary recommendation is to incorporate multiple maintenance strategies, at least one of which should facilitate transfer of control of the target behavior to other reinforcers. In addition, graded removal of the intervention, which has also been evaluated to a limited extent in CM, is a reasonable candidate for further development and evaluation. Direct comparisons of maintenance procedures are lacking, and should be considered a research priority in both domains. Researchers and clinicians interested in either type of intervention will likely benefit from ongoing attention to developments in both areas.

This research was supported by a grant from the National Institute on Drug Abuse (T32DA035200) of the National Institutes of Health. The funding agency had no role in study design, data collection or analysis, or preparation and submission of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This study was also funded in part by the Intramural Research Program of the NIH, National Institute on Drug Abuse (NIDA). Amanda Devoto is now at the American Society of Addiction Medicine, Rockville MD. The authors would like to thank Dr. David Jarmolowicz for his helpful comments on an early version of this manuscript.

Extended author information available on the last page of the article

Keywords Post-intervention outcomes · Relapse prevention · Generalization and maintenance · Incentive-based behavioral interventions · Substance use treatment

Introduction

Contingency management (CM) is typically construed as a category of consequence-based interventions used to change the frequency of health-related target behaviors. To be successful, these interventions require precise and objective measurement of behavior. They also typically involve provision of token reinforcers (i.e., cash, gift-cards, vouchers), exchangeable for goods and services, contingent on the occurrence (or nonoccurrence) of the target behavior. CM has been successfully applied to medication adherence (DeFulio & Silverman, 2011), physical activity and weight management (e.g., Kurti & Dallery, 2013), diabetes management (e.g., Raiff & Dallery, 2010), and breastfeeding (Washio et al., 2017). However, the dominant application of modern CM is to reduce the use of drugs, including opioids, stimulants, alcohol, cannabis, and nicotine (Davis et al., 2016; Lussier et al., 2006; Prendergast et al., 2006). CM interventions have been shown to be effective in well over 100 randomized controlled trials, and several meta-analyses (e.g., Dutra et al., 2008) have found them to be the most effective psychosocial intervention for substance use disorders. Despite this robust evidence base, adoption of CM in clinical settings has been slow, in part due to persistent criticisms.

Among the most common criticisms of CM is that the treatment effects of the intervention typically do not persist long after the intervention is removed. First, it should be noted that any behavior that is not reinforced will decline in frequency. This is a basic principle of operant conditioning. Second, substance use disorders are chronic conditions (McLellan et al., 2000). Thus, relapse is a common outcome after any substance use disorder treatment, and is not a special weakness of CM.

Posttreatment effects, also called treatment maintenance, are partly attributable to the process of stimulus generalization. Stimulus generalization is said to occur when a behavior persists in the presence of novel stimuli, most commonly an environment outside the one in which the target behavior was trained. This was most notably described by Stokes and Baer (1977), who emphasized the importance of programming interventions in a manner that would facilitate stimulus generalization. According to Stokes and Baer, generalization can be claimed when no further experimental manipulations are required to produce the desired changes in behavior. In the case of CM for substance use disorder, stimulus generalization could reasonably be incorporated into an account of treatment maintenance, because by definition the latter includes continuation of the behavior after the clinical supports that produced it have been withdrawn.

A full conceptual account of treatment maintenance will not be attempted here. However, we speculate that in addition to stimulus generalization, successful maintenance of drug abstinence is often a product of the establishment of instructional control of abstinence-related behaviors by rules, and more direct control of these behaviors from a variety of reinforcing consequences, at least some of which are likely social in nature. This account is consistent with the first study to show robust posttreatment effects in CM. That study featured an intensive behavioral

program called the community reinforcement approach (CRA) that involved rearrangement of the participants' work, family, and social environment to promote abstinence (Hunt & Azrin, 1973). Since the 1990s, CRA has been combined with voucher-based reinforcement therapy (VBRT), a CM program in which participants earn vouchers, exchangeable for goods and services, contingent upon the provision of a drug test indicating recent abstinence (Higgins et al., 1993). In a seminal study of CM treatment maintenance, Higgins et al. (2000) evaluated a CM intervention for people with cocaine use disorder. Participants were randomly assigned to the CRA plus vouchers contingent on cocaine abstinence, or a control group that received CRA with yoked noncontingent vouchers. Participants who received CRA with vouchers had significantly greater abstinence from cocaine over an 18-month follow-up period (Higgins et al., 2000). In a later study, Petry et al. (2017) found Higgins et al. to be 1 of 21 studies that showed a significant improvement in outcomes relative to controls months after the CM interventions were discontinued, and a meta-analysis found CM to be more effective than other evidence-based approaches in reducing objective indices of drug use up to a year after treatment (Ginley et al., 2021). Although these findings are promising, development of procedures that enhance posttreatment effects of CM intervention is still a research priority in the domain of substance-use disorder, because specific procedures that produce or at least promote treatment maintenance have not been identified, and substantial room for improvement remains.

The Token Economy

Another intervention that relies on the provision of token reinforcers contingent upon a target behavior is the token economy. As with most CM interventions, token economy interventions involve reinforcing desirable behavior with tokens that can be later exchanged for preferred items, privileges, and activities, called "back-up reinforcers." Token economies have a long history of clinical application, including modifying the behavior of individuals with mental illnesses in psychiatric facilities (Ayllon & Azrin, 1965), increasing attendance and task completion with children in juvenile court (Phillips et al., 1971), and promoting classroom participation (Boniecki & Moore, 2003). Token economies are easily disseminated and typically potent interventions. They have been implemented successfully across all manner of settings, with target populations and behaviors of all kinds.

Similarities between CM and Token Economies

Token economies involve similar procedures and the same behavioral processes as CM interventions. Most important, all token economies and the large majority of CM interventions produce their effects through the delivery of generalized conditioned reinforcers (i.e., conditioned reinforcers that have been paired with several primary reinforcers). For example, money is effectively established as a reinforcer by an enormous number of motivating operations, in that it can be exchanged for

anything that can be bought (Kelleher & Gollub, 1962; Skinner, 1953). In a token economy, a token may be exchangeable for many different back-up reinforcers and can maintain a target behavior in the face of changes in preferences for back-up reinforcers, whether temporary or permanent (Skinner, 1953). This results in less need to tailor reinforcers to specific individuals or conduct frequent preference assessments in the face of shifting motivational operations. Tokens also bridge the delay between a response and a terminal reinforcer (Kelleher, 1966; Skinner, 1953; Wolfe, 1936), which is especially useful in situations in which the delivery of the primary reinforcer is disruptive or impossible. In addition, incentives provided in CM and token economy interventions can be saved rather than spent immediately (Subramaniam et al., 2017), which opens access to reinforcers that would otherwise be too expensive to be practical.

Beyond procedural and conceptual similarities, treatment maintenance is a critical issue in both domains. The social significance of treatment maintenance was recognized early in the development of these interventions. For example, Winett and Winkler (1972) discussed importance of using behavior analytic techniques to lead independent, meaningful lives outside of treatment, rather than conforming to the status quo of their current environment. The purpose of Ayllon and Azrin's (1965) seminal work on token economies was to develop procedures that promoted independence for participants with mental health disorders. Likewise, Shedletsky and Voineskos (1976) discussed token economy systems used by inpatient psychiatric hospitals, which have often failed to adequately integrate patients into the community. The authors discuss a procedure for gradually fading out a token economy system as a participant is integrated back into the community and the importance of community-based programs to maintain treatment gains produced by the token economy.

Advantages of the Token Economy Literature as a Source of Postintervention Maintenance Strategies

Because of the procedural and conceptual similarities between CM and token economies, findings in token economy research are likely to be generalizable to CM interventions. As such, the token economy literature appears to be an ideal source for candidate procedures that could improve treatment maintenance outcomes in CM. It is important to note that the token economy literature has several features that make it a more fertile ground for candidate procedures than the CM literature itself. Chief among these is that CM literature has relatively few experiments dedicated to systematic analysis of specific design features of the intervention. Current CM procedures are essentially unchanged since the development of voucher-based (Higgins et al., 1991) and prized-based (Petry et al., 2000) CM. Most CM studies that involve evaluation of procedures investigate reinforcement magnitude and changes in reinforcement magnitude over time (e.g., Dallery et al., 2001; Roll et al., 2006; Roll & Higgins, 2000; Silverman et al., 1999), whereas the remainder are largely focused on aspects of drug testing (e.g., Chutuape et al., 2001; Correia et al., 2003; Correia et al., 2005). At present, a review of the CM literature on procedures for promoting postintervention maintenance is

premature due to a paucity of relevant research. In contrast, the token literature includes many studies designed to evaluate procedures within token economy interventions. A secondary advantage of reviewing the token economy literature is that studies in this research domain often feature single-subject designs. Single-subject designs are better suited for identifying critical intervention features and parameters (see Hackenberg, 2009, 2018, for substantial reviews on token economy procedures evaluated in single-subject designs and demonstrations of their use in explaining important behavioral processes). In contrast, the CM literature is comprised largely of group designs (see Petry et al., 2017; Dutra et al., 2008, for overviews of CM randomized controlled trials). Such designs are the gold standard for evaluating treatment effects at a group level, but they depend on inferential statistics and averaging of data, and as such obscure behavioral processes apparent when analysis is conducted at the level of the individual. Although some of the token economy studies included in the present review feature group designs, the more frequent use of single-subject designs in this literature is harmonious with the goals of this review. Overall, the combination of a much larger relevant token economy literature and the generalizability of token economy research to the domain of contingency management has establish the token economy as the strongest source of information regarding postintervention maintenance strategies for use in CM.

Maintenance Strategies

Treatment maintenance has been a topic of token economy research since the 1960s. Jones and Kazdin (1975) provided a brief overview of techniques that promote treatment maintenance after the discontinuation of a token economy intervention. As part of their introduction, they highlighted several procedures for promoting maintenance. First, stimuli previously paired with the delivery of token reinforcers, typically social reinforcers like peer and teacher praise, have been used after the token economy has been withdrawn (e.g., Chadwick & Day, 1971). Second, participants in a token program can be taught to implement the contingencies themselves (Fuoco et al., 1988). This is called self-monitoring, and removes the requirement of a teacher or caregiver to maintain the token contingency. In addition, intermittent reinforcement schedules have been used to promote resistance to extinction (Kazdin & Polster, 1973). To promote maintenance, token economies have also been faded prior to their complete removal (e.g., Sullivan & O'Leary, 1990). McIlvane and Dube (1992) describe fading as a procedure in which a stimulus is gradually removed as a method of transferring stimulus control. This can be achieved by manipulating the intensity, shape, or form of a stimulus, or the temporal relation between the onset of a prompt and stimulus (Schlichenmeyer et al., 2015). In a token economy this typically involves manipulating the physical properties of a token or gradually removing caregiver involvement in running the token economy. Though not discussed by Jones and Kazdin (1975), a final important maintenance procedure often described interchangeably with fading is called "thinning," which involves

gradually increasing one of the response requirements in a token economy. This can be accomplished by increasing the response requirement for earning a token (i.e., token production schedule), exchanging a token (i.e., exchange production schedule), or by increasing the requirement for producing a token exchange (i.e., token exchange schedule).

Decades of token maintenance research has been produced since the overview provided by Jones and Kazdin (1975). Despite the vast body of literature, token maintenance research has yet to be synthesized. The purpose of the present review is to synthesize and evaluate the token maintenance literature, and use the results to inform approaches to maintenance in CM. The present synthesis of the token maintenance literature, though undertaken to promote innovation in CM intervention design, could also be of value to a broader array of behavior analysts concerned with maintenance of the effects of token economy interventions.

Methods

Articles included in the review were identified using the databases Web of Science and PsycInfo. The first and third authors conducted all searches, and each reviewed the entirety of search results; accepting articles for further review or applying exclusion criteria. Figure 1 shows the search process and results.

Literature Search Procedures

The databases Web of Science and PsycInfo were searched. The term "Token*" was required to be in the title of the returned articles. In addition, at least one of the following Topic terms was required to be included: maint*, withdraw, extinction, intermittent, generaliz*, programmed, long-term, momentum, posttreatment, relapse, renewal,

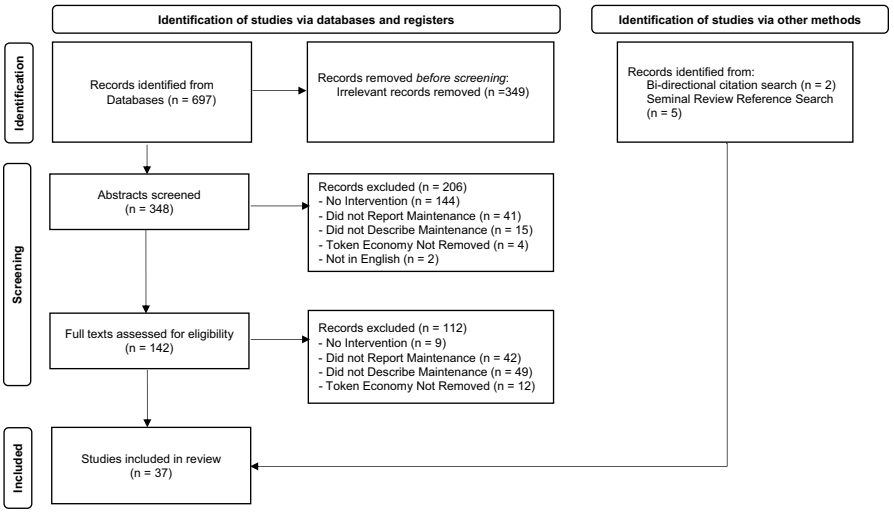

Fig. 1 Flowchart Illustrating the Literature Search Process and Results

resurgence, thin*, social-reinforc*, self*, social*, contingency management, follow-up. This latter criterion was implemented by placing the Boolean operator "OR" between each topic term. "Conditioned reinforcer" and "secondary reinforcer" were not included as terms because they generated a high level of off-topic items, and because of the ubiquity of the term "token" in the relevant literature.

In addition to the database searches any articles citing and cited by those included in the review based on the inclusion criteria below were reviewed in an iterated process until no further articles were obtained. Finally, the reference sections of four seminal reviews (Hackenberg, 2009, 2018; Kazdin, 1982; Kazdin & Bootzin, 1972), were searched for titles not returned by the prior steps in the literature search process.

Inclusion Criteria

To be included in the present review, studies were required to (1) be written in or translated to English; (2) be peer-reviewed; (3) evaluate a token economy intervention delivered to modify a target behavior; (4) report maintenance, return-to-baseline, or follow-up data; (5) describe methods used to engender maintenance of the effects of the token economy; and (6) remove the intervention as part of the study procedures. The first two criteria were incorporated into the database search by applying search filters.

Title Search

The initial Web of Science and PsycInfo searches were merged to remove duplicate titles, resulting in a list of 697 unique articles for initial review. An initial title search of those articles was conducted to remove articles that were not about token economies (e.g., tokenism, token populations, tokens in computer science). This resulted in the removal of 349 articles.

Abstract Review

The remaining 348 articles were subjected to abstract review by the first and third authors, who independently applied inclusion criteria. During the abstract review, inclusion criteria were applied in a rank order as described above with the article being removed upon a single criterion being unmet. Articles that that did not clearly violate any criteria were advanced for further consideration in the full text review. At this stage, a liberal process was adopted whereby any article judged as worthy of full review by either author was included in the full review. Among articles that were judged inappropriate for inclusion by both reviewers, the specific basis for exclusion was determined. Any disagreements were discussed by the first and third author and a final determination was made. Absence of an evaluation of a token economy intervention designed to modify target behavior resulted in the removal of 144 articles. Absence of any maintenance, reversal, or follow-up data resulted in the removal of 41 articles. A further 15 articles did not describe any systematic maintenance procedures. Four abstracts discussed follow-up but clearly described that the

token economy was never removed (e.g., were evaluations of long-term implementation of a token economy). Two articles were not in English. Thus, 142 articles advanced to full text review.

Full Text Review

Procedures for full review mimicked those used in abstract review. Initial IOA for the full-text review was 84.18% between the first and third authors. During the full text review, 42 articles were excluded for not reporting maintenance, follow-up, or a return to baseline. Forty-nine were excluded for not providing a description of maintenance procedures, and 12 articles did not remove the token economy during follow-up. Nine articles were removed for not using a token economy as part of a clinical intervention. Thus, the database search produced an initial body of 30 articles for review. The bidirectional iterated citation search of these 30 articles yielded an additional two articles for inclusion. Finally, an additional five articles were included based on the review of the reference sections of seminal reviews of the token economy literature by Kazdin or Hackenberg. This resulted in a final set of 37 articles included in the present review.

Article Categorization

Manuscripts accepted for inclusion in this review were examined and categorized by maintenance procedure. All authors participated in this process.

Thinning

Any articles were first categorized as "thinning" if the authors explicitly stated that the token economy was thinned and removed prior to the follow-up phase of the study, with or without the provision of procedural details. Second, if not explicitly stated, any articles in which at least one of the three token component schedules were thinned prior to follow-up were also included in the thinning category.

Fading

Articles were classified as fading if the authors stated that they used fading but did not provide additional procedural details. However, there were circumstances when the authors stated that fading was used but described a procedure in which a token component schedule was increased, which is a reinforcement schedule thinning procedure. In articles in which fading was used to describe a thinning procedure, this is likely because the schedule under which a token was produced or exchanged was considered a stimulus property. For the purposes of this review, fading was defined as procedures in which the physical characteristics of a token were manipulated or when a caregiver was trained to gradually decrease their role in a token economy's procedures. Based on this definition, articles in which maintenance procedures were

originally classified as "fading" were reclassified as "thinning" as needed (see Fox & Roseen, 1977).

Transfer of Control

Any articles that included a programmed attempt to transfer stimulus control to stimuli that were present in the natural environment (i.e., social reinforcers) or an alternative source of reinforcement (i.e., self-monitoring interventions) were classified as "transfer of control."

Combination

Articles were classified as "Combination" if methods from multiple maintenance strategies were incorporated prior to the follow-up phase of the study. Because social reinforcers naturally occur in the environment, only studies in which the authors explicitly state that social reinforcers were paired with token distribution in addition to another maintenance procedure were included in the combination category. Studies in which the authors instructed caregivers to provide praise as typical in addition to another maintenance procedure were categorized by the primary maintenance procedure.

Partial Retention of the Intervention

Finally, any articles in which the token economy was faded, or one of the component schedules were thinned, but the token economy was never completely removed, were categorized as "partial retention of the intervention."

Results

Overview

See Table 1 for a summary of the included articles. Of the 37 articles included, 10 (27%) were classified as thinning, 2 (5%) were classified as fading, 5 (14%) were classified as transfer of control, 16 (43%) used a combination of maintenance procedures, and 4 (11%) partially retained the token economy. Overall, unaltered token economies were implemented for a median of 3 weeks prior to the beginning of maintenance procedures and follow-up conditions were 4 weeks long. Though sample size varied substantially between classifications, articles classified as "Transfer of Control" (median = 6 weeks) and "Thinning" (median = 4 weeks) had the longest median implementation of an unaltered token economy and "Fading" (median = 1 weeks) and "Partial Retention" (median = 2 weeks) the shortest. Articles classified as "Fading" (median = 10 weeks) and "Not Entirely Removed" (median = 8 weeks) had the longest follow-up conditions, and "Thinning" (median = 3 weeks), the shortest.

Table 1 Summary of reviewed manuscripts

Maintenance Procedure	Source	Study *N*	Population	Target Behavior	Maintenance Program	Results	Condition Duration (Weeks)	
							Follow-Up	TE
Thinning	Butler (1979)	12	Female Chronic Schizophrenic Patients	Self-care, work, undesirable behavior	-	Target behaviors maintained during follow-up for less withdrawn patients	N/S	88
	Evans (1976)	1	8.5-year-old female	Trichotillomania	-	No hair pulling occurred during 12-week follow-up	12	14
	Fernandez et al. (1973)	11	Psychiatric in-patients	12 Individual and Group Activities	-	69.61% improvement maintained during follow-up	2	4
	Fox and Roseen (1977)	1	Toddler with Phenylketonuria	Medication Adherence		Adherence exceeded treatment levels during follow-up	26 & 52	3
	Kuypers et al. (1968)	6	Grade school children	Appropriate and inappropriate classroom behavior	-	Results did not maintain (token economy was not effective)	2	2.4
	McGinnis et al. (1999)	2	Middle School Boys	Classroom Work Pages	-	Results maintained during follow-up for one participant	2 Sessions	10 Sessions

Table 1 (continued)

Maintenance Procedure	Source	Study *N*	Population	Target Behavior	Maintenance Program	Results	Condition Duration (Weeks)	
							Follow-Up	TE
	Mottram et al. (2002)	3	7-year-old General Education Students	Disruptive Behavior	-	Percentage of disruptive behaviors maintained below levels observed during intervention	3	6
	Musser et al. (2001)	3	Children with social and behavioral disorders	Disruptive Behavior	-	Results maintained during follow-up for all participants	2	2
	Nay and Legum, L. (1976)	11	Special Education Middle School Students	Out of Seat and inappropriate verbal behavior	-	Target behaviors maintained at treatment levels during follow-up	2	N/S
	Nilsson (1976)	5	Children	Encopresis	-	Results maintained during follow-up for three participants	52	.86
Fading	Carton and Schweitzer (1996)	1	Child undergoing hemodialysis	Compliance	-	Results maintained in follow-up with minimal regression in compliance	12 & 18	9 Sessions
	Sullivan and O'Leary (1990)	10	Children with academic and behavior problems	Off-task behavior	-	Half of the participants maintained treatment gains during follow-up	1.9	1.14

Table 1 (continued)

Maintenance Procedure	Source	Study *N*	Population	Target Behavior	Maintenance Program	Results	Condition Duration (Weeks)	
							Follow-Up	TE
Natural Environment/Alternative Reinforcers	Chadwick and Day (1971)	30	Underachieving minority students	Academic achievement	All procedures terminated except social praise	Results maintained during follow-up for two of three target behaviors	2	6
	Elliott et al. (1979)	18	Institutionalized Males	Appearance, recreational activities, social interaction	Social Reinforcement	Social reinforcement group-maintained treatment gains	8	24
	Fuoco et al. (1988)	24	Residential psychiatric patients	Bedroom preparation	Self-monitoring	Results maintained during follow-up	1	1
	Reisinger (1972)	1	Anxiety-depressive female	Crying/smiling	Fading - compared tokens + social reinforcement vs social reinforcement only	No hospitalizations or further treatment referrals after discharge	4	6
	Shogren et al. (2011)	2	Children with Autism Spectrum Disorder	Appropriate classroom behavior	Self-monitoring	Results maintained at treatment levels during maintenance condition	8	5 Sessions
Combination of Maintenance Strategies	Aho (1978)	252	7^{th} & 8^{th} Grade Students	Disruptive Behavior	Thinning + Social Reinforcement	Statistically significant decrease in disruptions maintained during follow-up vs control	N/S	N/S

Table 1 (continued)

Maintenance Procedure	Source	Study *N*	Population	Target Behavior	Maintenance Program	Results	Condition Duration (Weeks)	
							Follow-Up	TE
	Barker et al. (1978)	10	Children with Developmental Disabilities	Daily Living Skills	Thinning + Social Reinforcement	Results maintained in follow-up with minimal deterioration in performance	16	3
	Campbell and Willis (1978)	32	Fifth-grade students	Fluency, flexibility, and elaboration in essay writing	Fading + Social Reinforcement	Results maintained during follow-up for two of three target behaviors	.57	3
	Cohen et al. (1979)	1	14-year-old boy with behavioral and learning disorders	Completing long division, on task behavior	Fading + self-monitoring	Maintenance during fading phase, decrease to 70% on task rate and 80% accuracy during self-recording phase	.86	1.7
	Hanel and Martin (1980)	8	Males with Intellectual Disabilities	Work Productivity	Self-administered Tokens + Goal Setting	Maintenance for four of six participants who were evaluated for maintenance	10–20 Sessions	28
	Hislop et al. (1973)	10	Children diagnosed "functionally retarded"	Appropriate classroom and academic behaviors	-	Improvements in target behavior during follow-up exceeded treatment phase	26	16

Table 1 (continued)

Maintenance Procedure	Source	Study *N*	Population	Target Behavior	Maintenance Program	Results	Condition Duration (Weeks)	
							Follow-Up	TE
	Jones and Kazdin (1975)	4	Children with developmental disabilities	Repetitive body movements	Teacher/peer praise + thinning	Low levels of inappropriate motor movements maintained during follow-up periods	12	2.1
	Kazdin and Polster (1973)	2	Adult males with developmental disabilities	Work productivity/social skills	Thinning (Intermittent Reinforcement) + Teacher Praise	Maintenance in performance for participant in intermittent reinforcement condition	5	5
	Mendham and Thorne (1984)	46	Maladjusted boys	Maturity and participation	Thinning + Self-monitoring (the latter received by only two participants)	Individual data per level not provided. General success in moving through the levels of the program and reduction of problem behavior	N/S	N/S
	Miller and Dermer (1979)	25	Patients with Schizophrenia	Social & daily living skills	Thinning + Social Reinforcement	Maintenance during follow-up	120	N/S

Table 1 (continued)

Maintenance Procedure	Source	Study *N*	Population	Target Behavior	Maintenance Program	Results	Condition Duration (Weeks)	
							Follow-Up	TE
	Novak and Hammond (1983)	28	Fourth grade classroom	Reading problem solving	Self-monitoring + descriptive praise	Self-monitoring and descriptive praise together maintained performance after token removal	1.9	1.9
	Rosenstein and Price (1994)	1	Male with Schizophrenia	Maladaptive Eating	Fading + Thinning + Praise	Pacing tape readministered without tokens due to increased eating speed; improvements maintained after tape discontinued	4	2
	Turkewitz et al. (1975)	8	Children with disruptive behavior	Disruptive Behavior	Fading + Self-monitoring	Disruptive behavior maintained at low rates during	.71	1.71
	Walker and Buckley (1972)	48	Grade school children	Academic achievement	Peer support, equating stimulus conditions, teacher training	Overall maintenance of target behavior. Peer programming and equating stimulus conditions had significantly greater maintenance than control group.	8	8

Table 1 (continued)

Maintenance Procedure	Source	Study *N*	Population	Target Behavior	Maintenance Program	Results	Condition Duration (Weeks)	
							Follow-Up	TE
	Walker et al. (1976); Experiment 1	5	Elementary school children	Appropriate classroom behavior	Fading + Social Praise	Results maintained with minimal deterioration in appropriate behavior	1.43	2
	Wood and Flynn (1978)	6	Pre-delinquent youths	Room cleaning	Fading + Self-Monitoring	Results better maintained after Self-monitoring conditions than external token conditions	3.14	2.71
Token Economy Not Entirely Removed	Frederiksen and Frederiksen (1975)	14	Developmental Disabilities	On-task and disruptive behavior	Self-administered tokens	Target behaviors did not deteriorate with self-administered tokens.	8	3
	Hersen et al. (1972)	27	Adults in Psychiatric Hospital	Work task completion, hygiene	Thinning	Productivity increased during thinning conditions	N/S	1.71
	Phillips et al. (1971); Experiment II	6	Pre-delinquent youths	Room cleaning	Fading	Results maintained across fading contingencies	24	1.86

Table 1 (continued)

Maintenance Procedure	Source	Study *N*	Population	Target Behavior	Maintenance Program	Results	Condition Duration (Weeks)	
							Follow-Up	TE
	Santogrossi et al. (1973)	9	Emotionally Disturbed Children	Disruptive Behavior	Self-administered tokens	Disruptive behavior returned to baseline levels during self-administration phases; large discrepancy between self-ratings and actual display of disruptive behavior	1.9	3.14

See Tables 2 and 3 for a summary of maintenance results following the removal of the token economy for the studies included in this review. Follow-up target behavior outcomes decreased by a median of 73% from baseline in studies in which the objective was to decrease a problem behavior (e.g., disruptive behavior). Change from baseline ranged from 86% (Fading; though only one study was included) to 46% (Partial Retention). Thinning, which comprised of over half of studies in which the objective was to decrease a target behavior, had a median percentage from baseline of 75%. For studies in which the objective was to *increase* a target behavior (e.g., on task), follow-up outcome measures increased by a median of 58% from baseline. The category "Combination of Maintenance Strategies" supplied 42% of all studies with this objective and had a median 130% from baseline values, nearly double the next category ("Thinning", 83%). "Transfer of Control" (33%) and "Fading" (31%) had the lowest increase from baseline. However, "Fading" had only one study representing studies with this objective.

Thinning

Thinning was the most common standalone procedure used to promote the maintenance of behavior change produced by token-based intervention. In a thinning procedure at least one of the token component schedules is gradually increased. These include the *token production schedule*, *exchange production schedule*, and *token exchange schedule*.

Token Production Schedule

Token production thinning is the most straightforward and the most common thinning procedure identified in this review. The token production schedule specifies the schedule requirement for earning tokens. For example, in a study incorporating a token economy to increase self-care and decrease undesirable behavior for women living with schizophrenia, the authors increased the number of consecutive mornings a participant would be required to wake up on time to earn their tokens (Butler, 1979). They found that improvements gained during treatment maintained throughout the thinning phase and follow-up. Similar results have been found with other target behaviors (e.g., hair pulling; Evans, 1976) and in in classroom settings (see McGinnis et al., 1999; Nay & Legum, 1976).

Exchange Production Schedule

The exchange production schedule is the schedule requirement for producing the token exchange period, which is typically a requirement that a specific number of tokens must be earned to produce the opportunity to exchange. The exchange production schedule occurs *after* a token has been earned, independent from the token production response requirement. Musser et al. (2001) and Mottram et al. (2002) describe similar token economies used to decrease disruptive behavior and the subsequent thinning of the exchange production schedule by increasing the number of tokens required to earn a random prize included in envelope in a classroom setting. In both studies, target behaviors remained at intervention levels for all participants in follow-up conditions lasting several weeks.

Table 2 Summary of Maintenance following the Removal of the Token Economy

Attempting to Increase or Decrease Target Behavior?	Maintenance Category	Source	Dependent Variable	Average Values at Each Condition			% Change from Baseline	Values Provided by Authors or Calculated?*
				Baseline	Treatment	Follow-Up		
Decrease	Natural Environment/Alternative Reinforcers	Reisinger (1972)	Crying Frequency	29	2	3	90%	Calculated
	Fading	Carton and Schweitzer (1996)	Intervals Noncompliant	7.3	0.22	1	86%	Provided
	Thinning	Mottram et al. (2002) P1	% Disruptive Behavior	54%	12%	9%	83%	Provided
	Thinning	Nilsson (1976)	Soiling Frequency	1.97	0.41	0.37	81%	Calculated
	Combination of Maintenance Strategies	Jones and Kazdin (1975)	% Inappropriate Motor Movement	62.20%	30%	12.30%	80%	Provided
	Thinning	Mottram et al. (2002) P3	% Disruptive Behavior	54%	14%	11%	80%	Provided
	Thinning	Musser et al. (2001) P1	% Disruptive Behavior	39%	9%	9%	77%	Provided
	Thinning	Mottram et al. (2002) P2	% Disruptive Behavior	59%	16%	16%	73%	Provided
	Token Economy Not Entirely Removed	Frederiksen and Frederiksen (1975)	Disruptions	0.39	0.21	0.11	72%	Calculated
	Thinning	Musser et al. (2001) P3	% Disruptive Behavior	36%	12%	11%	69%	Provided
	Combination of Maintenance Strategies	Hislop et al. (1973)	% Inappropriate Behavior	52%	40.27%	17.20%	67%	Provided

Table 2 (continued)

Attempting to Increase or Decrease Target Behavior?	Maintenance Category	Source	Dependent Variable	Average Values at Each Condition			% Change from Baseline	Values Provided by Authors or Calculated?*
				Baseline	Treatment	Follow-Up		
	Thinning	Musser et al. (2001) P2	% Disruptive Behavior	36%	9%	12%	67%	Provided
	Combination of Maintenance Strategies	Turkewitz et al. (1975)	Disruptive Behavior/20-s Interval	1.33	0.43	0.6	55%	Provided
	Natural Environment/Alternative Reinforcers	Elliott et al. (1979)	NOISE Score (Withdrawal)	23.9	18.9	14.6	39%	Provided
	Thinning	Kuypers et al. (1968)	% Problem Behavior	54%	27.80%	41.50%	23%	Provided
	Token Economy Not Entirely Removed	Santogrossi et al. (1973)	Disruptive Behavior/20-s Interval	1.5	0.3	1.2	20%	Calculated
	Thinning	Evans (1976)	Hair Pulls Frequency		2.14	0	No Baseline Data Provided	Calculated
Increase	Combination of Maintenance Strategies	Kazdin and Polster (1973)	Interactions/day	0.8	11	10.06	1157.50%	Provided
	Combination of Maintenance Strategies	Rosenstein and Price (1994)	Time to eat (minutes) on average	3	14.94	18.22	507.33%	Calculated
	Combination of Maintenance Strategies	Novak and Hammond (1983)	Reading problems completed (mean)	1.58	9.21	7.74	389.87%	Provided

Table 2 (continued)

Attempting to Increase or Decrease Target Behavior?	Maintenance Category	Source	Dependent Variable	Average Values at Each Condition			% Change from Baseline	Values Provided by Authors or Calculated?*
				Baseline	Treatment	Follow-Up		
	Natural Environment/Alternative Reinforcers	Chadwick and Day (1971)	Problem solved/min	1.4	3.35	5.5	292.86%	Provided
	Combination of Maintenance Strategies	Wood and Flynn (1978)	Items correct out of 15	3.35	12	12.55	275%	Provided
	Thinning	McGinnis et al. (1999)	Michael—time on math	27%	100%	100%	270.37%	Calculated
	Combination of Maintenance Strategies	Cohen et al. (1979)	Accuracy %	31%	65%	80%	158.06%	Provided
	Combination of Maintenance Strategies	Walker et al. (1976)	appropriate behavior %	47%	78%	95%	102.13%	Provided
	Thinning	Butler (1979)	Percentage Scores	44%	77%	84%	90.91%	Provided
	Token Economy Not Entirely Removed	Hersen et al. (1972)	# of points earned	10.39	16.64	18.77	80.65%	Calculated
	Thinning	Fox and Roseen (1977)	Mean oz. consumed	9.13	12.17	16	75.25%	Calculated
	Combination of Maintenance Strategies	Campbell and Willis (1978)	Test Scores	5.4	8.19	9.22	70.74%	Calculated
	Natural Environment/Alternative Reinforcers	Chadwick and Day (1971)	Problems correct %	50%	70%	73%	46%	Provided

Table 2 (continued)

Attempting to Increase or Decrease Target Behavior?	Maintenance Category	Source	Dependent Variable	Average Values at Each Condition			% Change from Baseline	Values Provided by Authors or Calculated?*
				Baseline	Treatment	Follow-Up		
	Combination of Maintenance Strategies	Barker et al. (1978)	Group success rate %	66.67%	89.65%	95.59%	43.38%	Calculated
	Natural Environment/Alternative Reinforcers	Shogren et al. (2011)	Appropriate intervals %	68.36%	90.50%	92.50%	35.31%	Provided
	Natural Environment/Alternative Reinforcers	Elliott et al. (1979)	Social Interest-NOISE Score	21.5	30	28.6	33.02%	Provided
	Token Economy Not Entirely Removed	Frederiksen and Frederiksen (1975)	On task %	73.25	94.17	96.63	31.92%	Calculated
	Fading	Sullivan & O'Leary (1990)	On task %	54.57%	79.60%	71.38%	30.80%	Calculated
	Combination of Maintenance Strategies	Walker and Buckley (1972)	appropriate behavior %	53.91%	90%	70.09%	30.01%	Provided
	Combination of Maintenance Strategies	Cohen et al. (1979)	On Task %	59.50%	92%	70%	17.65%	Provided
	Natural Environment/Alternative Reinforcers	Elliott et al. (1979)	Competence-NOISE Score	45.5	54.9	52.1	14.51%	Provided
	Natural Environment/Alternative Reinforcers	Fuoco et al. (1988)	Scores (percentage)	71.45	90	80.55	12.74%	Provided

Table 2 (continued)

Attempting to Increase or Decrease Target Behavior?	Maintenance Category	Source	Dependent Variable	Average Values at Each Condition			% Change from Baseline	Values Provided by Authors or Calculated?*
				Baseline	Treatment	Follow-Up		
	Natural Environment/Alternative Reinforcers	Chadwick and Day (1971)	At work %	39%	57%	42%	7.69%	Provided
	Thinning	McGinnis et al. (1999)	Charlie—Time spend on math	35.70%	92.20%	15%	-42.02%	Calculated
	Thinning	Fernandez et al. (1973)	Percent improvement	0	66.55%	69.60%	No Baseline Data Provided	Provided
	Natural Environment/Alternative Reinforcers	Reisinger (1972)	Smiling frequency	0	23	19	No Baseline Data Provided	Calculated
	Token Economy Not Entirely Removed	Phillips et al. (1971)	# of items completed		396	359.17	No Baseline Data Provided	Calculated

Summary of treatment maintenance across all studies, sorted by largest change from baseline at follow-up. Individual subject data was included for studies that did not average across participants. Data was not provided and therefore not calculable for Nay and Legum, L. (1976), Aho (1978), Hanel and Martin (1980), Mendham and Thorne (1984), and Miller and Dermer (1979)

*Data not provided by authors was hand-calculated using graphs included in the manuscript. Therefore, calculated data is subject to error.

Table 3 Summary of Quantitative Information by Maintenance Procedure

Maintenance Procedure	Condition Duration (Weeks)		Median % Change from Baseline
	Follow-up	TE	
Overall			
Thinning	2.5	3	-
Fading*	8.45	1.14	-
Transfer of Control	4	6	-
Combination	3.14	2.1	-
TE Never Removed	8	3	-
Decrease Target Behavior			
Thinning	2.5	2.2	75%
Fading*	15		86%
Transfer of Control*	6	15	65%
Combination	12	2.1	67%
TE Never Removed	4.95	3.07	46%
Increase Target Behavior			
Thinning*	20.5	4	83%
Fading*	1.9	1.14	31%
Transfer of Control	3	6	33%
Combination	2.52	2.36	130%
TE Never Removed	16	1.86	56%

*Procedures in this category had 2 or fewer articles used in the calculation. Interpret results with caution.

Exchange production schedule thinning can also occur by decreasing the availability of exchange periods. In one study, Fox and Roseen (1977) describe a procedure in which a child received tokens for adhering to Lofenalac, an infant powder formula used to treat phenylketonuria. At first, token exchange was available upon demand, but this was reduced over time to once per week. Adherence exceeded treatment levels during follow-up observations for one year after the intervention was discontinued.

Token Exchange Schedule

The token exchange schedule is the behavioral requirement for exchanging tokens. In an applied setting the token exchange schedule is most commonly the cost of the back-up reinforcers on the menu (see Hackenberg, 2018, for a detailed explanation of the elements of a token economy). In one study, the cost of the back-up reinforcers was increased in a token economy program designed to increase group and individual activities for individuals living in a psychiatric inpatient facility (Fernandez et al., 1973). The authors report an improvement from treatment maintained during follow-up. Similar results have been found in a token economy used to increase appropriate use of the bathroom, with results maintaining during a follow-up phase for three of the five participants (Nilsson, 1976).

Thinning Multiple Schedules

One study manipulated the exchange production and token exchange schedules. Kuypers et al. (1968) incorporated a token economy to increase appropriate classroom behavior. The exchange production schedule was increased by decreasing the frequency of exchange opportunities, whereas the token exchange schedule was manipulated by increasing the number of tokens required for prizes. The results did not maintain during the 2-week follow-up phase. However, the authors note that the initial token economy was only moderately effective, likely due to poor treatment integrity.

Fading

The term "fading" is used liberally in the token economy literature, but only one study met the present definition of fading. In a stimulus fading procedure the physical characteristics of a stimulus (e.g., size, shape, color, frequency of presentation or prompts) are manipulated. In practice, a clinician or caregiver may gradually decrease their role in a token economy's implementation. In the lone clear example of fading, Sullivan and O'Leary (1990) reduced the frequency of a teacher visually scanning a classroom to provide tokens to children that were on task from four to zero times per 20 min over 8 days. Follow-up assessment was conducted about 2 weeks later, and treatment gains were maintained for half of the participants. In another study that possibly constitutes fading, a token economy was applied to noncompliance for a child receiving hemodialysis (Carton & Schweitzer, 1996). The child received one token for every 30-min interval without noncompliance with hemodialysis. After decreasing noncompliance using the token economy, the authors stated that the token economy was faded over several weeks with treatment outcomes maintaining at a 3- and 6-month follow-up period. However, a lack of procedural details prohibit conclusive categorization of this study.

Transfer of Control

Social Reinforcers

Programmed delivery of social reinforcers is the most common approach for achieving transfer of stimulus control from tokens to naturally occurring stimuli in the relevant environment. In particular, caregiver praise was frequently used as the core of social reinforcement procedures. The general strategy entailed concurrent delivery of praise and supportive statements with a token economy over some specified period, after which the token economy is removed. This strategy has been implemented as a maintenance strategy in classrooms (Chadwick & Day, 1971) and mental health institutions (Elliott et al., 1979; Reisinger, 1972), with results maintaining for up to 4 weeks (Reisinger, 1972).

Self-Monitoring

Another procedure for transferring stimulus control to natural contingencies is self-monitoring. In self-monitoring procedures, participants are taught to evaluate their own performance. For example, in one study adults in a residential psychiatric program were trained to self-monitor the condition of their bedrooms using a checklist (Fuoco et al., 1988). Tokens were delivered contingent upon accurate checklist completion. During follow-up, tokens plus self-monitoring condition outperformed comparison conditions, including a combination of tokens and praise, and single-strategy interventions such as self-monitoring without tokens.

In addition to monitoring their progress, participants may be responsible for the delivery of incentives, a task typically assigned to research personnel or caregivers. This usually occurs after a specific criterion has been met and the caregivers have gradually removed their involvement in token delivery. For example, Shogren et al. (2011) taught kindergarten students to complete a tracking sheet which they used to monitor rule following during classroom activities. Positive marks (i.e., smiley faces) served as tokens exchangeable for preferred snacks. The self-monitoring intervention promoted appropriate classroom behavior during an 8-week maintenance condition during which teachers were instructed to use the checklists however they wished.

Combining Maintenance Procedures

Concurrent use of two or more procedures described above was the most common approach to promoting maintenance. Most often such these treatment packages featured either thinning or fading.

Thinning Plus another Procedure

Token schedule thinning has often been combined with the provision of social reinforcers. For example, Aho (1978) evaluated a maintenance procedure designed to decrease disruptive behavior in a group design study. Teachers were instructed to deliver social reinforcers (i.e., praise, approval) contingent upon appropriate classroom behavior during a token reinforcement and thinning phase. Following the thinning and subsequent removal of the token economy, students who received the maintenance procedure were significantly less likely to engage in target behavior relative to controls. Similar results were found in a program designed to promote the completion of daily living activities (Barker et al., 1978). This combined procedure has also been shown to reduce psychiatric rehospitalizations for recently discharged patients (Miller & Dermer, 1979). Peer praise, rather than caregiver praise, has also been used successfully in conjunction with thinning in a classroom setting (Jones & Kazdin, 1975)

Thinning has also been combined with self-monitoring. In one study, a 14-year-old boy earned tokens for remaining on task and accurately recording his performance in a notebook during math instruction (Cohen et al., 1979). The duration of on-task time required to earn a token was gradually increased over 5 days and then removed entirely for a 6-day self-monitoring maintenance condition. On-task

behavior and accuracy initially decreased relative to the thinning condition, but the latter recovered after 3 days. These results have been replicated with task completion and extended to decreasing inappropriate behavior (Mendham & Thorne, 1984).

A final thinning technique is to decrease the number of participants that are randomly selected to be evaluated and receive tokens. This has the effect of reducing the probability of reinforcement for all participants equally, and is typically used in combination with social praise. This group thinning strategy has been especially prevalent in classroom settings. For example, Campbell and Willis (1978) successfully used this strategy to promote improvements in essay writing for 32 fifth-grade students. In that study, token and praise combination thinning procedures were conducted across 12 days. These results have also been extended to other typically developing, elementary school aged children (Walker et al., 1976) and also children with developmental disabilities (Hislop et al.,1973).

Fading Plus another Procedure

Self-monitoring is also used in conjunction with fading. This strategy produced successful maintenance when used as part of a room cleaning intervention for adolescents in a residential rehabilitation program (Wood & Flynn, 1978), and when used in an intervention for decreasing disruptive behavior for children with academic and social problems (Turkewitz et al., 1975). To assist the transition to independent self-monitoring, some researchers have incorporated a matching phase, where selected participants only received points if their ratings were within one point of teacher ratings. To fade the matching criteria, the percent of participants selected for matching is gradually reduced, and participants provide their tokens independently (Turkewitz et al., 1975).

Other Combinations

A variety of unique combination procedures have been successful in promoting maintenance. Kazdin and Polster (1973) measured the effects of social praise in combination with an intermittent token production schedule. In their study, a participant's social behavior was reinforced with tokens on an intermittent schedule. Following the intervention phase, the token contingency was removed and a 5-week return to baseline phase began, with only social reinforcers in effect. The authors found that the effects of the token economy maintained throughout return to baseline for the participant compared to a participant whose social behavior received continuous reinforcement. Social reinforcement has also been combined with a thinning and fading procedure to teach a psychiatric patient to slow their rate of food intake in order to prevent choking (Rosenstein & Price, 1994). Tape-recorded pacing instructions and tokens were used to initiate the behavior and then thinned to shape meal duration from 5 to 15 min. The first step toward maintenance was the fading and ultimate removal of the pacing instructions. A week later tokens were replaced with praise. During a final 4-week follow-up with praise but no pacing instructions or token economy in place the problem behavior did not occur. Similar social reinforcement procedures were used in a group design study to successfully promote academic problem solving (Novak & Hammond, 1983). Likewise, Hanel and Martin (1980) used investigated

combination of self-monitoring, client-administered token economies, and goal setting. This approach was successful in maintaining work productivity in a vocational program for adults with developmental disabilities (Hanel & Martin, 1980).

Walker and Buckley (1972) investigated several maintenance strategies to promote academic achievement in grade school children, but did not isolate the effects of particular procedures. The combination approaches, which featured token reinforcement and praise, outperformed control conditions in which no maintenance procedures were conducted.

There were several articles in which token economies were partially removed. Frederiksen and Frederiksen (1975) and Santogrossi et al. (1973) describe self-administered token economies used to decrease disruptive behavior in a special education program and psychiatric hospital school, respectively. In both studies, tokens were initially distributed by teachers, who provided tokens to participants contingent upon them following a list of classroom rules. Before initiating the final self-monitoring procedure, Santogrossi et al. (1973) incorporated a matching procedure similar to others described in this review (e.g., Turkewitz et al., 1975). In contrast, Frederiksen and Frederiksen required teachers to ask students if they received a token at half-hour intervals and provide an additional token contingent upon saying yes. In the final phase of each study, participants were responsible for implementing their own token economy without teacher intervention. It is perhaps surprising that Santogrossi et al. (1973) reported inaccurate self-ratings and a return of disruptive behavior, whereas Frederiksen and Frederiksen reported successful maintenance.

In another study, Hersen et al. (1972) incorporated a token exchange schedule thinning procedure to increase work task completion and personal hygiene for 27 adults living in a psychiatric hospital. Following the initiation of a token program the cost of back-up reinforcers was doubled over a 2-week period. The authors report an increase in productivity during this period. Likewise, Phillips et al. (1971) describe an application of a token economy to promote room cleaning for six predelinquent children in Achievement Place, a community-based behavior modification program. In the second of two experiments described in this manuscript, children received points for meeting objective room cleaning criteria. After point fading was completed, the point contingency was applied to 8% of possible days for 6 months, with results maintaining during that period.

Discussion

Four categories of maintenance procedures were outlined in this review: (1) thinning; (2) fading; (3) transfer of control; and (4) partial retention of the intervention. Maintenance procedures typically fell into one or a combination of these categories, with most studies incorporating a combinatory approach. Most important, all procedural categories show substantial promise as strategies for promoting maintenance of behavior change produced by token economy interventions. When the goal was to decrease the target behavior, thinning appeared to be the most popular approach, but no category appears to be noticeably superior. When the goal was to increase behavior, combinatory approaches were most popular, and also appear to be the

most effective. We now turn to more detailed evaluation of each of the procedural categories prior to discussing their implications for CM.

Thinning

Articles that incorporated a thinning strategy were the most common of the standalone maintenance strategies and produced larger than average changes from baseline independent of the direction of the desired effect. This suggests that thinning may be an important component to promoting long-term treatment maintenance. However, there are important considerations that must be made before drawing conclusions on the efficacy of thinning as a standalone maintenance strategy. First, studies in which thinning was used as the only maintenance strategy had the shortest median follow-up periods. The popularity of these procedures may be a product of the ease with which they can be implemented. However, thinning does not involve systematic transfer of stimulus control from the token to another stimulus. Thinning as a prelude to an extinction procedure may delay regression to baseline performance but should not be expected to produce indefinite maintenance. As such, thinning procedures may be conceptualized as a form of "train and hope" as described by Stokes and Baer (1977). The hope is that the gradual removal of tokens will increase the probability of other reinforcers maintaining treatment gains. It is possible that such incidental transfer of control becomes more likely as the duration of exposure to the original token economy contingencies is extended. This would be true in any case in which there was a direct correlation between behavior change produced by the token economy and an alternative source of reinforcement. For example, token economies may produce patterns of behavior that access contingencies that did not contact behavior in its baseline state, even if a behavior analyst did not program these contingencies. Across the four intervention categories, thinning procedures had the second longest token economy condition durations. Whether extending treatment duration enhances the effects of thinning procedures is an empirical question worthy of further investigation on conceptual and practical grounds.

Although *token production* schedule thinning was the most common procedure, thinning the *exchange production* schedule instead may be a more effective approach because it allows for more immediate and more frequent token deliveries following the target behavior. Increasing the exchange production schedule does not affect the time between the occurrence of the target behavior and the delivery of the token. Instead, the time between earning a token and having the opportunity to exchange the token for a backup reinforcer is increased. This latter operation is less disruptive to the reinforcing effects of tokens when compared to the former (Hackenberg, 2018). Likewise, thinning the token exchange schedule does not decrease the frequency of token delivery. Of course, maintaining the token production schedule while decreasing the exchange production schedule entails a progressive increase in the cost of the reinforcers per exchange prior to the removal of the token economy. Thus, another open question with respect to thinning as a means of promoting treatment maintenance is whether thinning of token production schedules, token exchange schedules, or a combinatory approach is most effective.

Fading

After removing articles that did not meet our definition of fading, only two articles included fading as a standalone maintenance procedure. This makes it particularly difficult to draw conclusions on their efficacy. Having the largest median follow-up duration (8 weeks) and a study that produced larger than average decreases from baseline is confounded by the low sample size. Fading may share similar problems with thinning as no other stimuli are programmed into the participants environment. Despite these limitations, fading may be particularly important in school settings, where teachers are primarily responsible for the implementation of token economies. Gradually removing their involvement in the provision of tokens may be effectively combined with other natural sources of reinforcement, like self-monitoring, as the measuring of target behavior and provision of tokens can be gradually transferred from the teacher to the student.

Transfer of Control

The most dependable standalone maintenance procedure may involve a programmed transfer of control of target responses to stimuli that occur naturally in the participant's environment (Ayllon & Azrin, 1965). Social reinforcers such as praise and approval are among the most common uses of generalized condition reinforcement as a part of everyday life and clinical interventions. Interventions involving social reinforcers have been widely published for over 50 years (Allen et al., 1964; Zimmerman & Zimmerman, 1962) and continue to be a common component of behavior treatment plans for individuals with developmental disabilities and mental illnesses. The widespread availability of natural sources of reinforcement offers a solution to the limitations of using thinning and fading, discussed above. It is assumed that "train and hope" strategies like thinning and fading produce their maintenance effects because a natural source of reinforcement has acquired stimulus control.

As an alternative to social reinforcement, several articles included a self-monitoring component to serve a similar purpose. Self-monitoring has a practical advantage over programmed social reinforcers because participants record their own behavior and often manage the delivery of putative reinforcers. The self-monitoring procedures described in the articles reviewed here were successful in all but one study. Nevertheless, their efficacy has been called into question by authors who observed poor outcomes when implementing self-management strategies. For example, Santogrossi et al. (1973) reported a regression in treatment gains following the use of a student determined point system, and a lack of correspondence between student and teacher ratings. However, in a similar study conducted a year prior, students maintained treatment levels of disruptive behavior during an 8-day self-evaluation condition (Kaufman & O'Leary, 1972). An especially successful example is provided by Shogren et al. (2011), who used a self-monitoring procedure to maintain treatment gains for 8 weeks. Important considerations in self-monitoring procedures are the initial stages of the program when participants are learning to accurately record their target behavior. To alleviate issues during this period, participants can be

taught to match the instructor's target behavior recording and token delivery. Wood and Flynn (1978) discussed the significance of this matching procedure rather than the abrupt transfer from teacher to student administered token economy. The results of this review support their endorsement of the matching procedure, because it was included in several manuscripts that demonstrated treatment maintenance. However, given the overall success of self-monitoring procedures, it is difficult to determine which components result in its effectiveness. In addition, these maintenance procedures produced worse than average changes from baseline for both studies in which the objective was to increase the frequency of one target behavior and decrease the frequency of another target behavior.

Combination of Maintenance Strategies

Graded addition of naturalistic reinforcers in combination with graded subtraction of contrived reinforcers appears to be the surest way to transfer control to the former with minimal disruption in performance. Nevertheless, this approach is not universally successful. Articles that included a combination of maintenance strategies comprised nearly half of the articles included in the review. Token economies were implemented for a median of 3 weeks, and follow-up durations were 4 weeks long, which were equal to the overall medians for both metrics. Studies in this category also had a median change from baseline nearly triple the overall median for studies in which the objective was to increase a target behavior. However, these studies performed slightly below the overall median for studies in which the objective was to *decrease* a target behavior, though only three "combination" articles comprised this group. Although the larger quantity of articles that included a combination of strategies does not provide evidence that they are more effective, it does suggest that there are many combinations of maintenance strategies that can promote effective maintenance. Using a combination of maintenance procedures allows a practitioner to capitalize on the benefits of each one. For example, thinning, which produced overall higher than average changes from baseline during follow-up, may be combined with a natural source of reinforcement, such as social reinforcers. Of articles that included a combination of maintenance strategies, about 70% included a programmed social reinforcement component. This aligns with recommended strategies to promote response generalization, such as those described by Stokes and Baer (1977) who endorsed a programmed transfer of control to naturally occurring stimuli. Social reinforcement procedures are prevalent in the treatment of problem behavior and development of important skills. They are also often relatively easy to program in natural environments. Social reinforcers combined with thinning may extend the duration of maintenance beyond what was shown in the "Thinning" studies included in this review, which typically had the shortest follow-up periods.

CM Maintenance Literature

The CM literature can be fairly criticized as lacking in the use and evaluation of strategies for promoting postintervention maintenance of treatment effect. Nevertheless, it is noteworthy that thinning has occasionally been attempted in CM

research. In a recent meta-analysis of CM trials that reported drug use up to 1 year after the discontinuation of incentives, Ginley et al. (2021) found six articles that included fading. In that analysis, fading was defined as procedural modifications "where reinforcers are reduced or become more variable over time." This manipulation of schedules of reinforcement is better described as thinning according to the definitions offered in this review. In any case, the strategy appears to have been unsuccessful in increasing drug abstinence during follow-up (Ginley et al., 2021). Several parallels exist between thinning strategies in CM and token economies. For example, in a study investigating low-cost CM on illicit drug use, Alessi et al. (2008) used thinning by decreasing the number of names that were drawn out of a bowl in a group contingency. Participants whose name was drawn had a chance to receive prizes contingent upon providing a negative urine sample. This is like the procedure described by Campbell and Willis (1978), who decreased the number of students who were selected to receive tokens for improved essay writing.

In other studies the number of urine samples a participant submitted per week was decreased over the duration of the intervention (Dallery et al., 2015; Petry et al., 2006). Likewise, Andrade et al. (2014) evaluated a thinning procedure to promote maintenance of physical activity as measured by steps. The frequency of step monitoring was decreased over 12 weeks. These studies model methods used by Phillips et al. (1971), who decreased the percentage of days a token point contingency was applied over 6 months. Like Phillips et al., in both studies target behaviors improved significantly during treatment and throughout thinning procedures, but the differences were not maintained at follow-up. It is important to note that the 24-week follow-up phase used by Andrade et al. (2014) was significantly longer than the average duration of follow-up for token studies that used a thinning maintenance strategy (i.e., 3 weeks). It is possible and perhaps even likely that token economy interventions produce similar results over similar timeframes.

Implications for CM

The results of this review provide several considerations that are of direct relevance to the promotion of better long-term outcomes after CM intervention. The first step to programming treatment maintenance is the systematic removal of the monetary incentives that are delivered contingent upon a target behavior. The three component schedules that were typically thinned in the reviewed manuscripts (i.e., token production, exchange production, and exchange) also apply to CM. In a CM intervention, the token production schedule is the response requirement to produce an incentive. Targets such as medication adherence and attendance at counseling lend themselves especially well to manipulation of the token production schedule. For drug abstinence, duration is the relevant response parameter. Based on studies identified by Ginley et al. (2021) token production thinning appears to be the most common in CM. This includes decreasing the percentage of days participants are required to submit a urine sample (see Andrade et al., 2014; Petry et al., 2006). The exchange production schedule is another candidate for thinning and could be manipulated by increasing the minimum amount that needs to be earned before a payout is possible, or restricting payouts to certain days of the week or a limited number of times per week. Finally, In CM, a participant may exchange monetary incentives for items at a store, which have various prices. In

this case, a token exchange schedule manipulation would include fluctuating the price of the items. Note however, that cash-like incentives delivered via smart debit cards are becoming increasingly common in CM (DeFulio, 2022), and increase the convenience of the interventions while allowing participants to access a wider array of backup reinforcers.

Of the three component schedules, thinning the exchange production schedule offers several attractive features. First, unlike token production thinning, it stabilizes the overall magnitude of the monetary reinforcers during the intervention while still enhancing treatment maintenance. The general lack of a moderating effect of token production schedule thinning on follow-up abstinence observed by Ginley et al. (2021) suggests that increasing response requirements prior to earning an incentive is an ineffective approach to promoting maintenance. These procedures are more appropriate if the incentives are never removed entirely, as in Andrade et al. (2014). Second, unlike exchange schedules, increasing exchange production schedules does not require reducing the motivational conditions under which the incentives would effectively reinforce behavior, and does not disrupt the convenience afforded by smart debit cards. As an example of exchange production schedule thinning, a participant may earn $5 for each urine sample that indicates recent drug abstinence, with samples required three times per week on average. During initial phases of the program, the participant would have immediate access to their earnings, regardless of the amount saved. To thin the exchange production schedule, a minimum balance would be required prior to transfer of funds to the debit card. In a prize-based CM arrangement it may be appropriate to increase the number of negative samples a participant needs to provide to produce a draw from a prize bowl. Therefore, participants who submit three samples per week can still produce three prize bowl drawings. However, they can only draw from the bowl after earning three drawings. In a voucher-based program, participants would only be able to exchange their vouchers after earning a specified amount. This requirement could be increased over time and would allow for the gradual removal of the incentives without decreasing their value over time.

The robust effects of studies that used a combination of maintenance strategies suggest that in addition to the gradual removal of a CM program, programming alternative sources of reinforcement for the target behavior is important to promote maintenance. Using this strategy may significantly improve the CM thinning procedures. For example, developing healthy and supportive social relationships with people who do not use drugs can help facilitate sustained drug abstinence in people with substance use disorders (Pettersen et al., 2019) and social reinforcers have promoted abstinence and long-term aftercare attendance following intensive treatment for substance use disorders (Lash et al., 2004).

In addition to programmed social reinforcement, self-monitoring may play an important role in CM. Self-monitoring is a widely studied standalone approach for substance use, perhaps because of its low cost and low provider burden, but the results are mixed (Gass et al., 2021). Rather than implementing it as a standalone intervention, CM interventionists could consider modeling token self-monitoring research by integrating it with CM. Self-monitoring in CM may involve allowing participants to assess their own toxicology results and gradually provide their own incentives based on this objective measure. The matching procedure discussed by Wood and Flynn (1978) could further bolster CM self-monitoring and is also straightforward to fade over time.

Recently, technological innovations have increased the accessibility and scalability of CM (For reviews, see Dallery et al., 2019; Kurti et al., 2016). With technology, CM can be integrated with complementary evidence-based practices, such as cognitive behavior therapy (CBT; Epstein et al., 2003). CBT and other evidenced based practices such as Community Reinforcement and Family Training (CRAFT) can help facilitate the crucial transfer of stimulus control to naturally occurring stimuli in a participant's environment. Use of at least some of the maintenance strategies discussed in this review are essential to the utility smartphone-based CM platforms. In addition to the exchange production schedule thinning, certain features of smartphone-based CM interventions could be systematically faded out. For example, appointment reminders may become less frequent and less salient over time. This may be modeled after token studies that reduced the frequency of caregiver involvement (i.e., Sullivan & O'Leary, 1990). In addition, drug tests and their notifications may become less frequent as participants reach abstinence goals. Finally, selected elements of the platform (e.g., drug testing, medication reminders, CBT modules) could remain on the platform as the incentives are removed, and eventually removed in turn.

As described in the introduction, another operant approach to promoting drug abstinence is the community reinforcement approach (CRA). CRA includes explicit attempts to program natural sources of reinforcement in addition to the use of incentives. According to Hunt and Azrin (1973) those receiving CRA are assisted by a counselor in finding meaningful employment, and a lawyer if legal problems are present. Those who struggle with employment are provided vocational counseling (e.g., building a resume, developing interview skills). Participants also receive marital and family counseling that focused on the incompatibility between substance use and healthy romantic, family, and social relationships. Finally, participants are assisted in getting access to necessities such as a telephone and public transportation, which would give them access to powerful social reinforcers. Participants who receive CRA have higher rates of abstinence, are more often gainfully employed, and have better relationships with their families (see Higgins et al., 2003; Hunt & Azrin, 1973). In more recent programs, CRA has incorporated CM to initiate and sustain drug abstinence (Higgins et al., 2003; Secades-Villa et al., 2008). Although intensive, this program has many features that correspond with the results of this review such as using access to less restrictive environments (e.g., Hislop et al., 1973; Reisinger, 1972) and social reinforcers from peers and caregivers (e.g., Elliott et al., 1979) to maintain behavior after material reinforcers have been withdrawn.

Generalization Challenges

There are several challenges to generalizing the results of this review to populations who typically receive CM for substance use disorders. Many of these challenges stem from a lack of experimental control over other reinforcers. For example, several studies in this review that included social reinforcers as a maintenance strategy programmed a specific number of social praises participants would receive throughout a specified time frame (e.g., Walker et al., 1976). These rates of face-to-face

social reinforcement delivery are more feasible in an environment where caregivers and participants have several hours of interactions each day. In addition, a procedure such as the one used by Jones and Kazdin (1975), in which a student stands in front of their classmates to receive applause, does not seem feasible in the context of CM, except in special settings that feature group counseling. However, a CM practitioner should be able to program social reinforcers through some mechanism that is easy to access for everyone involved (e.g., peer recovery coaches, support groups, drug-abstinent friends). As an alternative, or perhaps supplementary to social reinforcement, the token reinforcement literature suggests that self-monitoring has promise as a method for achieving transfer of control in CM intervention.

Another potential difference between the studies included in this review and the current procedures used in most CM interventions is a difference in available backup reinforcers. In the token literature there is typically a limited number of backup reinforcers. In contrast, modern CM procedures often involve the use of cash-like incentives delivered via reloadable gift card (DeFulio, 2022). The latter approach provides access to a nearly unlimited number of backup reinforcers. Future research on the menu options influence outcomes is warranted, especially in the context of token economies.

Gaps in the Literature

Because only two included articles described maintenance failures, it is difficult to isolate specific variables that lead to the success or failure of a maintenance procedure. Therefore, additional studies are necessary to compare the effects of different maintenance procedures and other procedural details like the length of follow-up conditions, the duration of the intervention phase, the rate in which fading or thinning occurs, and the use of other natural reinforcers besides social praise.

Comparison of Maintenance Procedures

Significant procedural differences between the studies included in this review make the comparison of maintenance procedures difficult. Few studies compared two or more included maintenance procedures. Comparisons were typically between a maintenance procedure and a control (e.g., Aho, 1978; Kazdin & Polster, 1973). The most directly relevant study compared posttreatment maintenance of self-monitoring, descriptive feedback plus praise, and a combination of self-administration and praise (Novak & Hammond, 1983). This is a largely understudied area and warrants further investigation to promote best practice maintenance procedures. In addition, the metric used to compare studies in this review (i.e., percent change from baseline*)* should be interpreted cautiously due to the variability in settings, target behaviors, and participant characteristics across studies. In addition, the number of examples of studies within each procedural category was imbalanced across categories. Finally, studies in which the objective was to decrease a target behavior could not deviate more than 100% from baseline, as a target behavior cannot be below zero. However, there is no ceiling on percent from

baseline for studies with the objective of increasing a target behavior. For example, Kazdin and Polster (1973) increased the number of interactions per day from 0.8 per day to 10.06 per day, a 1157.50% change from baseline. This makes it difficult to compare studies in which the goal was to decrease the rate of a behavior to studies in which the goal was to increase the rate of a behavior.

Length of Follow-Up

The length of follow-up conditions varied substantially across articles included in this review. Despite this heterogeneity, the length of the follow-up condition appears to be a critical variable when evaluating the effectiveness of a maintenance procedure. It is possible that longer follow-up phases would have detected eventual relapse of problem behavior. This is especially true for the studies that implemented a single maintenance procedure, like thinning, which typically had a shorter follow-up period (median of 2 weeks), excluding a 1-year outlier (Nilsson, 1976). Articles that included multiple maintenance strategies had about double the follow-up period (median of 4 weeks). This inconsistency makes comparing maintenance procedures difficult as one procedure may appear more successful due to having a shorter follow-up condition. Further, it is ideal to measure maintenance of behavior change over months or years, not weeks. Although long-term follow-up is often time consuming and expensive, it is essential to advance our understanding of the conditions that lead to meaningful maintenance of treatment gains.

Duration of the Intervention

Abstinence during treatment and follow-up attendance often increase as the length of CM implementation increases (Roll et al., 2013). However, these effects are variable (see Lussier et al., 2006; Prendergast et al., 2006; Roll et al., 2013) and are predominately not extended to treatment maintenance. Although there are several meta-analyses measuring the effects of CM duration on abstinence, this is a largely understudied topic in token economy research, especially related to maintenance following the removal of a token economy intervention. The length of the token economy phases of the included studies were variable. However, differences were noted. Santogrossi et al. (1973) highlight the 9-day token economy phase prior to its removal compared to other studies that implemented the token program for over 3 weeks (e.g., Frederiksen & Frederiksen, 1975; Kaufman & O'Leary, 1972), which may have resulted in early relapse. Despite these observations, there is currently no research on of the effects of the length of token economy implementation on postintervention maintenance.

Thinning/Fading Rate

Another understudied variable that may play a large role in treatment maintenance is the rate in which a token economy is thinned or faded prior to its removal. The effects of thinning and fading rates on treatment maintenance have yet to be

studied. The only related topic that has been researched is ratio strain. It is typically recommended to gradually increase token production schedules to promote treatment effects during the thinning process (Roane et al., 2007). However, there are no empirically derived recommendations on the rate to thin or fade a token economy. This is an important detail for treatment maintenance. Increasing the length of the thinning/fading phase of an intervention increases the duration the patient has exposure to the contingencies generated by the intervention. This is related to other important maintenance procedures, such as the pace in which alternative reinforcers are simultaneously programmed into the patient's environment.

Transfer of Control

A majority of studies included in this review that incorporated a natural reinforcement maintenance procedure used social praise and approval by caregivers (e.g., Aho, 1978; Barker et al., 1978; Elliott et al., 1979). However, naturalistic reinforcers may be more substantially integrated into a token economy program to promote postintervention maintenance. For example, rather than reminding caregivers to provide a remark such as "that was good" contingent upon a desirable response (see Campbell & Willis, 1978), a more powerful social contingency may be applied. Peer-managed token economies are equally effective compared to caregiver managed (Bedell & Archer, 1980) and may provide a patient with more frequent access to powerful, peer-mediated social reinforcers. Incorporating peer praise into a token economy program successfully promoted treatment maintenance (Jones and Kazdin (1975). However, this has never been compared to the effects of social reinforcement provided by a caregiver (see Elliott et al., 1979), and maintenance following peer-mediated token programs has not been assessed.

Treatment progress often results in access to natural sources of reinforcement that were unavailable due to problematic behavior. For example, Hislop et al. (1973) stated that participants who completed the intervention were able to transition from a remedial to a typical classroom setting. However, this was a biproduct of meeting a treatment goal; the authors did not assess the effects of access to a general education setting as a maintenance strategy. This setting may have provided an increased access to important natural sources of reinforcement, thus promoting maintenance. In addition, access to new settings may translate to CM treatment maintenance, as sobriety may increase access to important natural sources of reinforcement, like improved family and social relationships (Pettersen et al., 2019).

Frequency of Maintenance Studies over Time

Researchers have long recognized that maintenance of treatment effects is essential to meaningful change for individuals who receive behavioral intervention services (e.g., Shedletsky & Voineskos, 1976; Winett & Winkler, 1972) As shown in this review, the interest in treatment maintenance grew significantly from the 1970s to the 1990s. However, only three articles included in this review were published after the year 2000 (Mottram et al., 2002; Musser et al., 2001;

Shogren et al., 2011). This paucity of recent research is troubling given the continued popularity of token reinforcement procedures in practice and the many unanswered questions and the central importance of treatment maintenance. However, CM is an incredibly active current area of research, thus it may be possible to conduct the necessary studies in that domain. Such research could and should then inform token economy intervention design.

Conclusions

Thinning, fading, and programmed transfer of control can and should be implemented to promote treatment maintenance in CM and other contingency-based interventions in behavior analysis. The development of smartphone-based CM (e.g., DeFulio et al., 2021) has created a context in which doing so has become more practical than ever. CM researchers and behavior analysts more broadly should vigorously build on the extensive but still incomplete literature describing the conditions under which the effects of token-based reinforcement interventions are maintained after the interventions are discontinued.

References

Aho, S. (1978). Modification of disruptive behavior: The adaptation of token reinforcement, model learning and role playing methods to overcome disruptions of orderly working conditions at school. *Scandinavian Journal of Educational Research, 22*(2), 49–64.

Alessi, S. M., Petry, N. M., & Urso, J. (2008). Contingency management promotes smoking reductions in residential substance abuse patients. *Journal of Applied Behavior Analysis, 41*(4), 617–622. https://doi.org/10.1901/jaba.2008.41-617

Allen, K. E., Hart, B., Buell, J. S., Harris, F. T., & Wolf, M. M. (1964). Effects of social reinforcement on isolate behavior of a nursery school child. *Child Development, 35*(2), 511–518.

Andrade, L. F., Barry, D., Litt, M. D., & Petry, N. M. (2014). Maintaining high activity levels in sedentary adults with a reinforcement-thinning schedule. *Journal of Applied Behavior Analysis, 47*(3), 523–536.

Ayllon, T., & Azrin, N. H. (1965). The measurement and reinforcement of behavior of psychotics. *Journal of the Experimental Analysis of Behavior, 8*(6), 357–383. https://doi.org/10.1901/jeab.1965.8-357

Barker, P., Docherty, P., Hird, J., & Hunter, M. H. (1978). Living and learning: A nurse administered token economy programme involving mentally handicapped schoolboys. *International Journal of Nursing Studies, 15*(2), 91–102. https://doi.org/10.1016/0020-7489(78)90020-2

Bedell, J. R., & Archer, R. P. (1980). Peer managed token economies: Evaluation and description. *Journal of Clinical Psychology, 36*(3), 716–722. https://doi.org/10.1002/1097-4679(198007)36:3

Boniecki, K. A., & Moore, S. (2003). Breaking the silence: Using a token economy to reinforce classroom participation. *Teaching of Psychology, 30*(3), 224–227. https://doi.org/10.1207/S15328023TOP3003_05

Butler, R. J. (1979). An analysis of individual treatment on a token economy for chronic schizophrenic patients. *British Journal of Medical Psychology, 52*(3), 235–243. https://doi.org/10.1111/j.2044-8341.1979.tb02520.x

Campbell, J. A., & Willis, J. (1978). Modifying components of "creative behavior" in the natural environment. *Behavior Modification, 2*(4), 549–564. https://doi.org/10.1177/014544557824007

Carton, J. S., & Schweitzer, J. B. (1996). Use of a token economy to increase compliance during hemodialysis. *Journal of Applied Behavior Analysis, 29*(1), 111–113. https://doi.org/10.1901/jaba.1996.29-111

Chadwick, B. A., & Day, R. C. (1971). Systematic reinforcement: Academic performance of underachieving students. *Journal of Applied Behavior Analysis, 4*(4), 311–319. https://doi.org/10.1901/jaba.1971.4-311

Chutuape, M. A., Silverman, K., & Stitzer, M. L. (2001). Effects of urine testing frequency on outcome in a methadone take-home contingency program. *Drug & Alcohol Dependence, 62*(1), 69–76. https://doi.org/10.1016/s0376-8716(00)00160-5

Cohen, R., Rubin, S., & Heinen, J. R. (1979). Collateral increases in performance of long division resulting from on-task gains using token reinforcement and self-recording. *Psychological Reports, 44*(2). https://doi.org/10.2466/pr0.1979.44.2.651

Correia, C. J., Dallery, J., Katz, E. C., Silverman, K., Bigelow, G., & Stitzer, M. L. (2003). Single- versus dual-drug target: Effects in a brief abstinence incentive procedure. *Experimental & Clinical Psychopharmacology, 11*(4), 302–308. https://doi.org/10.1037/1064-1297.11.4.302

Correia, C. J., Sigmon, S. C., Silverman, K., Bigelow, G., & Stitzer, M. L. (2005). A comparison of voucher-delivery schedules for the initiation of cocaine abstinence. *Experimental & Clinical Psychopharmacology, 13*(3), 253–258. https://doi.org/10.1037/1064-1297.13.3.253

Dallery, J., Meredith, S., Jarvis, B., & Nuzzo, P. A. (2015). Internet-based group contingency management to promote smoking abstinence. *Experimental & Clinical Psychopharmacology, 23*(3), 176–183. https://doi.org/10.1037/pha0000013

Dallery, J., Raiff, B. R., Grabinski, M. J., & Marsch, L. A. (2019). Technology-based contingency management in the treatment of substance-use disorders. *Perspective on Behavior Science, 42*(3), 445–464. https://doi.org/10.1007/s40614-019-00214-1

Dallery, J., Silverman, K., Chutuape, M. A., Bigelow, G. E., & Stitzer, M. L. (2001). Voucher-based reinforcement of opiate plus cocaine abstinence in treatment-resistant methadone patients: Effects of reinforcer magnitude. *Experimental & Clinical Psychopharmacology, 9*(3), 317–325. https://doi.org/10.1037//1064-1297.9.3.317

Davis, D. R., Kurti, A. N., Skelly, J. M., Redner, R., White, T. J., & Higgins, S. T. (2016). A review of the literature on contingency management in the treatment of substance use disorders, 2009–2014. *Preventive Medicine, 92*, 36–46. https://doi.org/10.1016/j.ypmed.2016.08.008

DeFulio, A. (2022). Dissemination of contingency management for the treatment of opioid use disorder. *Perspectives on Behavior Science*. Advance online publication. https://doi.org/10.1007/s40614-022-00328-z

DeFulio, A., & Silverman, K. (2011). Employment-based abstinence reinforcement as a maintenance intervention for the treatment of cocaine dependence: Post-intervention outcomes. *Addiction, 106*(5), 960–967. https://doi.org/10.1111/j.1360-0443.2011.03364.x

DeFulio, A., Devoto, A., Traxler, H., Cosottile, D., Fingerhood, M., Nuzzo, P., & Dallery, J. (2021). Smartphone-based incentives for promoting adherence to antiretroviral therapy: A randomized controlled trial. *Preventive Medicine Reports, 21*, Article 101318. https://doi.org/10.1016/j.pmedr.2021.101318

Dutra, L., Stathopoulou, G., Basden, S. L., Leyro, T. M., Powers, M. B., & Otto, M. W. (2008). A meta-analytic review of psychosocial interventions for substance use disorders. *American Journal of Psychiatry, 165*(2), 179–187. https://doi.org/10.1176/appi.ajp.2007.06111851

Elliott, P. A., Barlow, F., Hooper, A., & Kingerlee, P. E. (1979). Maintaining patients' improvements in a token economy. *Behaviour Research and Therapy, 17*(4), 355–367. https://doi.org/10.1016/0005-7967(79)90007-x

Epstein, D. H., Hawkins, W. E., Covi, L., Umbricht, A., & Preston, K. L. (2003). Cognitive-behavioral therapy plus contingency management for cocaine use: Findings during treatment and across 12-month follow-up. *Psychology of Addictive Behaviors, 17*(1), 73–82. https://doi.org/10.1037/0893-164x.17.1.73

Evans, B. (1976). A case of trichotillomania in a child treated in a home token program. *Journal of Behavior Therapy & Experimental Psychiatry, 7*(2), 197–198.

Fernandez, J., Fischer, I., & Ryan, E. (1973). The token economy: A living-learning environment. *British Journal of Psychiatry, 122*(569), 453–455. https://doi.org/10.1192/bjp.122.4.453

Fox, R. A., & Roseen, D. L. (1977). Parent administered token program for dietary regulation of phenylketonuria. *Journal of Behavior Therapy & Experimental Psychiatry, 8*(4), 441–443. https://doi.org/10.1016/0005-7916(77)90018-0

Frederiksen, L. W., & Frederiksen, C. B. (1975). Teacher-determined and self-determined token reinforcement in a special education classroom. *Behavior Therapy, 6*(3), 310–314.

Fuoco, F. J., Lawrence, P. S., & Vernon, J. B. (1988). Post-reinforcement effects of token reinforcement, verbal praise, and self-monitoring in a residential psychiatric program. *Behavioral Residential Treatment, 3*(4), 267–286. https://doi.org/10.1002/bin.2360030404

Gass, J. C., Funderburk, J. S., Shepardson, R., Kosiba, J. D., Rodriguez, L., & Maisto, S. A. (2021). The use and impact of self-monitoring on substance use outcomes: A descriptive systematic review. *Substance Abuse, 1–36*. https://doi.org/10.1080/08897077.2021.1874595

Ginley, M. K., Pfund, R. A., Rash, C. J., & Zajac, K. (2021). Long-term efficacy of contingency management treatment based on objective indicators of abstinence from illicit substance use up to 1 year following treatment: A meta-analysis. *Journal of Consulting & Clinical Psychology, 89*(1), 58–71. https://doi.org/10.1037/ccp0000552

Hackenberg, T. D. (2009). Token reinforcement: A review and analysis. *Journal of the Experimental Analysis of Behavior, 91*(2), 257–286. https://doi.org/10.1901/jeab.2009.91-257

Hackenberg, T. D. (2018). Token reinforcement: Translational research and application. *Journal of Applied Behavior Analysis, 51*(2), 393–435. https://doi.org/10.1002/jaba.439

Hanel, F., & Martin, G. (1980). Self-monitoring, self-administration of token reinforcement, and goal-setting to improve work rates with retarded clients. *International Journal of Rehabilitation Research, 3*(4), 505–517. https://doi.org/10.1097/00004356-198012000-00004

Hersen, M., Eisler, R. M., Smith, B. S., & Agras, W. S. (1972). A token reinforcement ward for young psychiatric patients. *American Journal of Psychiatry, 129*(2), 228–233. https://doi.org/10.1176/ajp.129.2.228

Higgins, S. T., Badger, G. J., & Budney, A. J. (2000). Initial abstinence and success in achieving longer term cocaine abstinence. *Experimental & Clinical Psychopharmacology, 8*(3), 377–386. https://doi.org/10.1037//1064-1297.8.3.377

Higgins, S. T., Budney, A. J., Bickel, W. K., Hughes, J. R., Foerg, F., & Badger, G. (1993). Achieving cocaine abstinence with a behavioral approach. *American Journal of Psychiatry, 150*(5), 763–769. https://doi.org/10.1176/ajp.150.5.763

Higgins, S. T., Delaney, D. D., Budney, A. J., Bickel, W. K., Hughes, J. R., Foerg, F., & Fenwick, J. W. (1991). A behavioral approach to achieving initial cocaine abstinence. *American Journal of Psychiatry, 148*(9), 1218–1224. https://doi.org/10.1176/ajp.148.9.1218

Higgins, S. T., Sigmon, S. C., Wong, C. J., Heil, S. H., Badger, G. J., Donham, R., Dantona, R. L., & Anthony, S. (2003). Community reinforcement therapy for cocaine-dependent outpatients. *Archives of General Psychiatry, 60*(10), 1043–1052. https://doi.org/10.1001/archpsyc.60.9.1043

Hislop, M. W., Moore, C., & Stanish, B. (1973). Remedial classroom programming: Long-term transfer effects from a token economy system. *Mental Retardation, 11*(2), 18–20 https://www.ncbi.nlm.nih.gov/pubmed/4786288

Hunt, G. M., & Azrin, N. H. (1973). A community-reinforcement approach to alcoholism. *Behaviour Research and Therapy, 11*(1), 91–104. https://doi.org/10.1016/0005-7967(73)90072-7

Jones, R. T., & Kazdin, A. E. (1975). Programming response maintenance after withdrawing token reinforcement. *Behavior Therapy, 6*(2), 153–164.

Kaufman, K. F., & O'Leary, K. D. (1972). Reward, cost, and self-evaluation procedures for disruptive adolescents in a psychiatric hospital school. *Journal of Applied Behavior Analysis, 5*(3), 293–309. https://doi.org/10.1901/jaba.1972.5-293

Kazdin, A. E. (1982). The token economy: A decade later. *Journal of Applied Behavior Analysis, 15*(3), 431–445.

Kazdin, A. E., & Bootzin, R. R. (1972). The token economy: An evaluative review. *Journal of Applied Behavior Analysis, 5*(3), 343–372.

Kazdin, A. E., & Polster, R. (1973). Intermittent token reinforcement and response maintenance in extinction. *Behavior Therapy, 4*(3), 386–391.

Kelleher, R. T. (1966). Conditioned reinforcement in second-order schedules. *Journal of the Experimental Analysis of Behavior, 9*(5), 475–485.

Kelleher, R. T., & Gollub, L. R. (1962). A review of positive conditioned reinforcement. *Journal of the Experimental Analysis of Behavior, 5*(4), 543–597.

Kurti, A. N., & Dallery, J. (2013). Internet-based contingency management increases walking in sedentary adults. *Journal of Applied Behavior Analysis, 46*(3), 568–581. https://doi.org/10.1002/jaba.58

Kurti, A. N., Davis, D., Redner, R., Jarvis, B., Zvorsky, I., Keith, D. R., Bolivar, H., White, T. J., Rippberger, P., Markeish, C., Atwood, G., & Higgins, S. T. (2016). A review of the literature on remote monitoring technology in incentive-based interventions for health-related behavior change. *Translational Issues in Psychological Science, 2*(2), 128–152. https://doi.org/10.1037/tps0000067

Kuypers, D. S., Becker, W. C., & O'Leary, K. D. (1968). How to make a token system fail. *Exceptional Children, 35*(2), 101–109.

Lash, S. J., Burden, J. L., Monteleone, B. R., & Lehmann, L. P. (2004). Social reinforcement of substance abuse treatment aftercare participation: Impact on outcome. *Addictive Behaviors, 29*(2), 337–342. https://doi.org/10.1016/j.addbeh.2003.08.008

Lussier, J. P., Heil, S. H., Mongeon, J. A., Badger, G. J., & Higgins, S. T. (2006). A meta-analysis of voucher-based reinforcement therapy for substance use disorders. *Addiction, 101*(2), 192–203. https://doi.org/10.1111/j.1360-0443.2006.01311.x

McGinnis, J. C., Friman, P. C., & Carlyon, W. D. (1999). The effect of token rewards on "intrinsic" motivation for doing math. *Journal of Applied Behavior Analysis, 32*(3), 375–379. https://doi.org/10.1901/jaba.1999.32-375

McIlvane, W. J., & Dube, W. V. (1992). Stimulus control shaping and stimulus control topographies. *The Behavior Analyst, 15*(1), 89–94. https://doi.org/10.1007/BF03392591

McLellan, A. T., Lewis, D. C., O'Brien, C. P., & Kleber, H. D. (2000). Drug dependence, a chronic medical illness: Implications for treatment, insurance, and outcomes evaluation. *JAMA, 284*(13), 1689–1695. https://doi.org/10.1001/jama.284.13.1689

Mendham, R. P., & Thorne, M. T. (1984). A description and evaluation of a "levelled token economy" operating within the school day in a residential school for junior maladjusted boys. *Behavioural Psychotherapy, 12*(2), 151–162. https://doi.org/10.1017/S0141347300009824

Miller, H. R., & Dermer, S. W. (1979). Quasi-experimental follow-up of token-economy and conventional treatment graduates. *Journal of Consulting & Clinical Psychology, 47*(3), 625–627 https://www.ncbi.nlm.nih.gov/pubmed/528735

Mottram, A. M., Bray, M. A., Kehle, T. J., Broudy, M., & Jenson, W. R. (2002). A classroom-based intervention to reduce disruptive behaviors. *Journal of Applied School Psychology, 19*(1), 65–74. https://doi.org/10.1300/J008v19n01_05

Musser, E. H., Bray, M. A., Kehle, T. J., & Jenson, W. R. (2001). Reducing disruptive behaviors in students with serious emotional disturbance. *School Psychology Review, 30*(2), 294–304.

Nay, R. W., & Legum, L. (1976). Increasing generalization in a token program for adolescent retardates: A methodological comment. *Behavior Therapy, 7*(3), 413–414.

Nilsson, D. E. (1976). Treatment of encopresis: Token economy. *Journal of Pediatric Psychology, 4*(1), 42–46.

Novak, G., & Hammond, J. M. (1983). Self-reinforcement and descriptive praise in maintaining token economy reading performance. *Journal of Educational Research, 73*(3), 186–189. https://doi.org/10.1080/00220671.1983.10885447

Petry, N. M., Alessi, S. M., Carroll, K. M., Hanson, T., MacKinnon, S., Rounsaville, B., & Sierra, S. (2006). Contingency management treatments: Reinforcing abstinence versus adherence with goal-related activities. *Journal of Consulting & Clinical Psychology, 74*(3), 592–601. https://doi.org/10.1037/0022-006X.74.3.592

Petry, N. M., Alessi, S. M., Olmstead, T. A., Rash, C. J., & Zajac, K. (2017). Contingency management treatment for substance use disorders: How far has it come, and where does it need to go? *Psychology of Addictive Behaviors, 31*(8), 897–906. https://doi.org/10.1037/adb0000287

Petry, N. M., Martin, B., Cooney, J. L., & Kranzler, H. R. (2000). Give them prizes, and they will come: Contingency management for treatment of alcohol dependence. *Journal of Consulting & Clinical Psychology, 68*(2), 250–257. https://doi.org/10.1037//0022-006x.68.2.250

Pettersen, H., Landheim, A., Skeie, I., Biong, S., Brodahl, M., Oute, J., & Davidson, L. (2019). How social relationships influence substance use disorder recovery: A collaborative narrative study. *Substance Abuse: Research & Treatment, 13*. https://doi.org/10.1177/1178221819833379

Phillips, E. L., Phillips, E. A., Fixsen, D. L., & Wolf, M. M. (1971). Achievement place: Modification of the behaviors of pre-delinquent boys within a token economy. *Journal of Applied Behavior Analysis, 4*(1), 45–59.

Prendergast, M., Podus, D., Finney, J., Greenwell, L., & Roll, J. (2006). Contingency management for treatment of substance use disorders: A meta-analysis. *Addiction, 101*(11), 1546–1560. https://doi.org/10.1111/j.1360-0443.2006.01581.x

Raiff, B. R., & Dallery, J. (2010). Internet-based contingency management to improve adherence with blood glucose testing recommendations for teens with type 1 diabetes. *Journal of Applied Behavior Analysis, 43*(3), 487–491. https://doi.org/10.1901/jaba.2010.43-487

Reisinger, J. J. (1972). The treatment of "anxiety-depression" via positive reinforcement and response cost. *Journal of Applied Behavior Analysis, 5*(2), 125–130. https://doi.org/10.1901/jaba.1972.5-125

Roane, H. S., Falcomata, T. S., & Fisher, W. W. (2007). Applying the behavioral economics principle of unit price to DRO schedule thinning. *Journal of Applied Behavior Analysis, 40*(3), 529–534. https://doi.org/10.1901/jaba.2007.40-529

Roll, J. M., Chudzynski, J., Cameron, J. M., Howell, D. N., & McPherson, S. (2013). Duration effects in contingency management treatment of methamphetamine disorders. *Addictive Behaviors, 38*(9), 2455–2462. https://doi.org/10.1016/j.addbeh.2013.03.018

Roll, J. M., & Higgins, S. T. (2000). A within-subject comparison of three different schedules of reinforcement of drug abstinence using cigarette smoking as an exemplar. *Drug & Alchohol Dependence, 58*(1–2), 103–109. https://doi.org/10.1016/s0376-8716(99)00073-3

Roll, J., Huber, A., Sodano, R., Chudzynski, J. E., Moynier, E., & Shoptaw, S. (2006). A comparison of five reinforcement schedules for use in contingency management-based treatment of methamphetamine abuse. *The Psychological Record, 56*, 67–81.

Rosenstein, L. D., & Price, R. F. (1994). Shaping a normal rate of eating using audiotaped pacing in conjunction with a token economy. *Neuropsychological Rehabilitation, 4*(4), 387–398. https://doi.org/10.1080/09602019408401607

Santogrossi, D. A., O'Leary, K. D., Romanczyk, R. G., & Kaufman, K. F. (1973). Self-evaluation by adolescents in a psychiatric hospital school token program. *Journal of Applied Behavior Analysis, 6*(2), 277–287. https://doi.org/10.1901/jaba.1973.6-277

Schlichenmeyer, K. J., Dube, W. V., & Vargas-Irwin, M. (2015). Stimulus fading and response elaboration in differential reinforcement for alternative behavior. *Behavioral Interventions, 30*(1), 51–64. https://doi.org/10.1002/bin.1402

Secades-Villa, R., García-Rodríguez, O., Higgins, S. T., Fernández-Hermida, J. R., & Carballo, J. L. (2008). Community reinforcement approach plus vouchers for cocaine dependence in a community setting in Spain: Six-month outcomes. *Journal of Substance Abuse Treatment, 34*(2), 202–207. https://doi.org/10.1016/j.jsat.2007.03.006

Shedletsky, R., & Voineskos, G. (1976). The rehabilitation of the chronic psychiatric patient: Beyond the hospital-based token economy system. *Social Psychiatry, 11*, 145–150. https://doi.org/10.1007/BF00579313

Shogren, K. A., Lang, R., Machalicek, W., Rispoli, M. J., & O'Reilly, M. (2011). Self- versus teacher management of behavior for elementary school students with Asperger syndrome: Impact on classroom behavior. *Journal of Positive Behavior Interventions, 12*(2), 87–96. https://doi.org/10.1177/1098300710384508

Silverman, K., Chutuape, M. A., Bigelow, G. E., & Stitzer, M. L. (1999). Voucher-based reinforcement of cocaine abstinence in treatment-resistant methadone patients: Effects of reinforcement magnitude. *Psychopharmacology, 146*(2), 128–138. https://doi.org/10.1007/s002130051098

Skinner, B. F. (1953). *Science and human behavior.* Macmillan.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*(2), 349–367. https://doi.org/10.1901/jaba.1977.10-349

Subramaniam, S., DeFulio, A., Jarvis, B. P., Holtyn, A. F., & Silverman, K. (2017). Earning, spending, and drug use in a therapeutic workplace. *The Psychological Record, 67*(2), 273–283. https://doi.org/10.1007/s40732-017-0237-0

Sullivan, M. A., & O'Leary, S. G. (1990). Maintenance following reward and cost token programs. *Behavior Therapy, 21*, 139–149.

Turkewitz, H., O'Leary, K. D., & Ironsmith, M. (1975). Generalization and maintenance of appropriate behavior through self-control. *Journal of Consulting & Clinical Psychology, 43*(4), 577–583. https://doi.org/10.1037/h0076905

Walker, H. M., & Buckley, N. K. (1972). Programming generalization and maintenance of treatment effects across time and across settings. *Journal of Applied Behavior Analysis, 5*(3), 209–224. https://doi.org/10.1901/jaba.1972.5-209

Walker, H. M., Hops, H., & Fiegenbaum, E. (1976). Deviant classroom behavior as a function of combinations of social and token reinforcement and cost contingency. *Behavior Therapy, 7*(1), 76–88. https://doi.org/10.1016/s0005-7894(76)80221-3

Washio, Y., Humphreys, M., Colchado, E., Sierra-Ortiz, M., Zhang, Z., Collins, B. N., Kilby, L. M., Chapman, D. J., Higgins, S. T., & Kirby, K. C. (2017). Incentive-based intervention to maintain breastfeeding among low-income Puerto Rican mothers. *Pediatrics, 139*(3). https://doi.org/10.1542/peds.2016-3119

Winett, R. A., & Winkler, R. C. (1972). Current behavior modification in the classroom: Be still, be quiet, be docile. *Journal of Applied Behavior Analysis, 5*(4), 499–504. https://doi.org/10.1901/jaba.1972.5-499

Wolfe, J. B. (1936). Effectiveness of token rewards for chimpanzees. *Comparative Psychology Monographs, 12*, 72.

Wood, R., & Flynn, J. M. (1978). A self-evaluation token system versus an external evaluation token system alone in a residential setting with predelinquent youth. *Journal of Applied Behavior Analysis, 11*(4), 503–512. https://doi.org/10.1901/jaba.1978.11-503

Zimmerman, E. H., & Zimmerman, J. (1962). The alteration of behavior in a special classroom situation. *Journal of the Experimental Analysis of Behavior, 5*(1), 59–60.

Authors and Affiliations

Sean D. Regnier[1] · Haily K. Traxler[1] · Amanda Devoto[2] · Anthony DeFulio[3]

✉ Anthony DeFulio
anthony.defulio@wmich.edu

[1] University of Kentucky College of Medicine, Lexington, KY, USA

[2] National Institute on Drug Abuse Intramural Research Program, Baltimore, MD, USA

[3] Department of Psychology, Western Michigan University, 1903 West Michigan Avenue, Kalamazoo, MI 49008, USA