Skeptic's Corner: Punishment — Destructive Force or Valuable Social “Adhesive”?

Thomas S. Critchfield

Published online: 16 April 2014

© Association for Behavior Analysis International 2014

Abstract Skepticism that normally focuses on pseudoscientific claims can also be directed at “established” principles of behavior. After discussing some ways in which empirically-derived principles can potentially mislead, as an illustrative example I describe some reasons to wonder whether our understanding of punishment is as established as sometimes assumed.

Keywords Punishment · Translation · Cooperation

Too often we hold fast to the clichés of our forebears. We subject all facts to a prefabricated set of interpretations. We enjoy the comfort of opinion without the discomfort of thought.
— John F. Kennedy¹

Preamble: Skepticism in Behavior Analysis

In 2008, *Behavior Analysis in Practice* launched an occasional feature called the “Skeptic’s Corner” that has been devoted to critical evaluation of pseudoscientific claims about behavior and its remediation. As Normand (2008) wrote at the time “In science, being skeptical does not mean doubting the validity of everything. Rather, to be skeptical is to judge the validity of a claim based on objective empirical evidence” (p.

42). Past contributions to the Skeptic’s Corner have examined the efficacy of prism lenses (Chok et al. 2010) and hyperbaric oxygen chambers (Lerman et al. 2008) in the treatment of problems associated with autism; the validity of claims that vaccines cause autism (Ahearn 2010); and the proliferation of psychopharmacological interventions for various behavior problems (van Haaren 2009; Wyatt 2009).

Skeptical evaluation of such claims is essential, because humans tend to adopt causal beliefs, including those about the factors that cause and remediate problem behavior, rapidly and uncritically on a gut-feeling level (Adcock 1995), and the flashes of “insight” through which causal inferences arise can be accompanied by powerful emotions and feelings of certainty (e.g., Metcalfe and Weibe 1987). Because “insight” tends to bypass critical thinking, clinical insights from outside of behavior analysis have spawned a wide range of practices that are at best ineffective and at worst directly harmful (e.g., Critchfield et al. *in press*; Normand 2008).

Skepticism, as Normand (2008) defined it, is healthy and essential in dealing with faddish, non-behavioral approaches, but there is no a priori reason to limit its application to incautious insights of purely clinical origin or to notions that arose outside of behavior analysis. Although behavior analysis is an empirical enterprise in which we are accustomed to thinking in terms of evidence, it is the nature of science to shift attention from empirical evidence to broad generalizations. For example, many empirical “facts” document the validity of the concept of positive reinforcement. Thanks to B.F. Skinner (e.g., 1938) and his many intellectual descendants, numerous empirical observations have been distilled into general theoretical principles, and generations of behavior analysts have been educated about positive reinforcement principles without necessarily encountering all of the evidence that supports them.

From a Skeptic’s perspective, then, we walk a very fine line in a maturing science like behavior analysis

¹ Yale University commencement speech, June 11, 1962; retrieved from <http://millercenter.org/president/speeches/detail/3370>.

that has gained confidence in the validity of its general principles. On the one hand, some ideas receive enough evidentiary support that we feel little compulsion to routinely double-check them. On the other hand, as G.K. Chesterton (1923) once observed, “There are only two kinds of people, those who accept dogmas and know it, and those who accept dogmas and don’t know it” (p. 86).

Dogma means “principles set down by an authority as incontrovertibly true” (<http://www.oxforddictionaries.com>), and thus is not so different from well-established scientific principles. Scientific principles must be derived and proposed by someone (the theorist as “authority”). Once they receive enough empirical support they may, as in the case of positive reinforcement, be regarded as “incontrovertibly true.” In the abstract, there is an obvious distinction between scientific principles and religious dogmas. Dogmas are embraced because of faith in the authority that endorsed them, whereas scientific principles are embraced because of the evidence that led an “authority” to propose them (Normand 2008). But in practice, is the distinction really so clear-cut?

In a discipline like behavior analysis, in which most professionals are not professional scientists (e.g., Critchfield 2011), many people learn about scientific principles from secondary sources without mastering the full corpus of primary evidence that supports them. For example, read Skinner’s classic conceptual works like *Science and Human Behavior* (1953), *Beyond Freedom and Dignity* (1971), and *Verbal Behavior* (1957) and you will find few references to primary empirical evidence. Standard texts on applied behavior analysis introduce the same principles but often say little about their empirical origins. Functionally if not intentionally, each new generation of behavior analysis must trust in the “authority” of those who have distilled our central principles from primary empirical evidence and of those who disseminate them.

To reiterate a point, such trust is not necessarily misplaced: We trust our “second-hand” principles because they are regularly put to the test, and found worthy, through ongoing research and successful practical application (Normand 2008). Yet contemporary “dogmas” about behavior – the things we accept as true without referring directly to evidence — *could* be erroneous or incomplete if there is some sort of disconnect between our beliefs and the evidence that supports them. Such a disconnect can arise in at least two ways.

One potential problem might be called translational overconfidence. It is customary in behavior analysis to interpret the everyday world by extrapolating from findings of laboratory research; this was Skinner’s (e.g., 1953) approach in his classic conceptual writings. Laboratory analyses that inspire

such interpretations often are so carefully controlled (e.g., Sidman 1960) that they produce robust effects whose internal validity is rarely in doubt (it is no accident that, unlike in some other parts of psychology, findings in the experimental analysis of behavior rarely are overturned). But the fact that our experimental research holds up well over time is not enough to verify that the findings have generality to the world outside of the laboratory. For instance, effects seen in the laboratory could be robust under tightly controlled conditions but trivial amidst the multivariable chaos of the everyday world, or factors that give rise to these effects might be manipulated in the laboratory in greater potency than can be expected to occur in the everyday world.

A second potential problem arises when generalizations about behavior are derived without considering all of the available evidence. Behavior analysts usually are careful to scrutinize evidence produced by behavior analysts, but what if there is relevant evidence from outside behavior analysis that has not been incorporated into our principles? Behavior analysis has been criticized as insular (e.g., Krantz 1972) and not without reason. Our community has always been uneasy about its connections to mainstream psychology (e.g., Malott et al. 2002; Ulman 1993), and Skinner (1993) himself observed, at the end of his career, that “We have been accused of building our own ghetto, of refusing to make contact with other kinds of Psychology. Rather than break out of the ghetto, I think that we should strengthen its walls” (p. 5). Institutions such as the Association for Behavior Analysis International, which hosts the world’s largest behavior analytic conference, and the Society for the Experimental Analysis of Behavior, which publishes two of our most prominent scientific journals, arose partly as an objection to the culture of mainstream psychology (e.g., Malott et al. 2002). Perhaps as a result, there is evidence that we do not often digest work arising outside of our “ghetto.” For instance, behavior analytic journals rarely cite non-behavior-analytic sources (Critchfield and Reed 2004).

A Case in Point: Punishment

If our own basic studies do not mean what we think they do, or if there is compelling behavioral evidence of which we are not aware, then it is at least possible that we have been too uncritical about some aspects of what we think we know about behavior. In the remainder of this essay, I discuss as a possible case in point one of the sacred cows of behavior analytic orthodoxy, the notion that punishment as a means of behavior control is harmful. My goal is to illustrate the application to behavior principles of the same skeptical filter that we apply to non-behavior-analytic work, but before proceeding, I must be absolutely clear about two things. First, this is a selective analysis in which I will address only certain aspects of

the conventional wisdom about punishment.² A thorough conversation regarding what we think we know about punishment is far beyond the scope of the present article, although I hope this article will contribute to our field's collective motivation to engage in that conversation. Second, I will not offer specific recommendations for clinical practice. Applied behavior analysts have developed effective interventions based on positive reinforcement and for the time being it makes sense to continue using what works. My present goal is simply to introduce an intriguing wrinkle to the necessary conversation about punishment by describing research from outside of behavior analysis that has the potential to change the way we think about the role of punishment in therapy and society.

Core Truths

Every student of behavior analyst encounters two Core Truths about punishment. The First Core Truth is a theoretical statement depicting punishment as a central component of operant learning. Punishment is a consequence that reduces the future frequency of behaviors upon which it is contingent (e.g., Catania 2007). This function of punishment has been recognized formally since at least Thorndike's (1898) pioneering research and has been supported by decades of basic research (e.g., Azrin & Holz, Dinsmoor 1998).³

The Second Core Truth is a practical consideration indicating that applications of punishment in the everyday world are fraught with unpleasant costs. According to conventional wisdom,

A variety of side effects and problems are often correlated with applications of punishment, including the elicitation of undesirable emotional responses and

aggression, escape and avoidance, and an increased rate of the problem behavior under nonpunishment conditions.... Other problems include... overusing punishment because of the negative reinforcement it provides for the punishing agent's behavior. (Cooper et al. 2007, pp. 336–337)

Broad-scale implications of punishment's side effects have been addressed colorfully by Sidman (2000) and, especially, Skinner in *Beyond Freedom and Dignity* (1971) and *Walden Two* (1948), in which he imagined the benefits of a world arranged without the destructive influence of aversive control. I assume that readers are familiar with traditional concerns about punishment because these are echoed in a wide variety of contemporary applied behavior analysis resources (e.g., Daniels and Daniels 2006; Pryor 2002; Sarafino 2012; Vargas 2009; Wheeler and Richey 2014) and because knowledge of them is part of applied behavior analysis certification requirements (<http://www.bacb.com>).

Behavior Analytic Evidence

According to the First Core Truth, punishment is half of Nature's arsenal of consequences. From this theoretical perspective, avoiding punishment in clinical settings is a bit like sending a fighter into the boxing ring with one arm tied behind his back. Presumably, clinicians need all of the tools of behavior change that they can get, so much depends on whether it is really true that, in the management of unwanted behavior, "the disadvantages far outweigh the benefits of [punishment] procedures" (Wheeler and Richey 2014, p. 363).

Despite its wide acceptance among behavior analysts, the Second Core Truth is subject to wider interpretation than the first. A key point of departure in this discussion is that our understanding of punishment is rooted in research and theory that are many decades old (e.g., for general reviews of the basic punishment literature, see Azrin and Holz 1966; Dinsmoor 1998). While there is nothing intrinsically problematic about old science — for instance, Boyle's 1662 law relating gas pressure to gas volume holds up just fine today — certain aspects of the history of punishment research may leave us inadequately informed about how punishment operates in everyday environments.

To be specific about the problem, note that seminal research on punishment traces mainly to the 1940s through the early 1960s, after which new punishment experiments became fairly rare (e.g., Critchfield and Rassmussen 2007). The leading formal theories of punishment were proposed very early on, with one-factor theory tracing to Thorndike (e.g., 1913) and two-factor theory tracing at least to Mowrer (1940), with a possible nod to Thorndike (1932). Even "recent" updates to

² My comments also are selective in that I will steer clear of certain practical matters regarding punishment's position in society. Examples include legal restrictions on clinical applications of punishment in many jurisdictions and the view of clinical punishment as unkind, abusive, or otherwise objectionable in many codes of clinical ethics. Lengthy treatises could be devoted to examining how punishment got to be so unpopular and how current legal and ethical stances square with the empirical evidence on punishment. In the latter case, interested readers may find that Cipani (2004) provides a good start.

³ A common corollary to the First Core Truth is that efforts to harness punishment for therapeutic purposes are destined to fail (e.g., Vargas 2009). This is a complex topic that deserves more attention than be provided here, but I will mention two relevant issues before moving on. First, many interventions seek to build new repertoires and, unlike positive reinforcement, punishment does not teach what to do. I agree that under most circumstances positive reinforcement is preferred for constructing new repertoires. Second, for a variety of reasons, it is often said that punishment interventions do not actually eliminate unwanted behavior. This claim runs counter to a number of published reports of effective applications of punishment, both alone and in combination with positive reinforcement procedures (e.g., Lindscheid, Iwata, Ricketts, Williams, & Griffiths, 1990; see also Lerman and Vorndran 2002).

these theories are many decades old (e.g., Deluty 1976; de Villiers 1980; Dinsmoor 1954; Rachlin and Herrnstein 1969).

Overall, most punishment research and theory development trace to before the birth of applied behavior analysis in roughly the 1960s and 1970s (see Rutherford 2009). From this, it can be inferred that most available studies of punishment were not designed with translational questions in mind. That is, they were inspired mainly by theoretical questions, and procedures used to answer them were devised in service of good experimental control rather than as a means of replicating key features of everyday environment. Typically, unconditioned punishers (like electric shock) were applied fairly intensely and frequently to the behavior of nonhuman subjects, under a limited range of conditions involving concurrent reinforcement of target and alternative behaviors (e.g., Azrin and Holz 1966). It is worth wondering, therefore, about the extent to which the conditions under which key punishment side effects were observed in the laboratory are replicated in the world outside the laboratory.

To be clear, it would be naïve to question the external validity of basic research simply because it occurs in an “artificial” laboratory environment. Good experimental control, which laboratories specialize in creating, is a prerequisite to external validity (i.e., one must understand clearly how variables operate within an experiment to have any chance of extrapolating beyond it). Yet good experimental control does not assure external validity. Cipani (2004) examined some of the ways in which laboratory punishment experiments differ from everyday circumstances and concluded that too little is known to reliably predict when punishment’s adverse effects will occur outside the laboratory. More generally, after surveying the behavior analytic literature on punishment, Lerman and Vorndran (2002) argued that:

Current knowledge about basic processes is insufficient for translation to application. The basic literature on some important relations remains incomplete.... More important, the extent to which findings with nonhumans and response-contingent electric shock can be extrapolated into the treatment of behavior disorders in clinical populations may be substantially restricted.... Research on punishment only tentatively supports most prescriptions for application. (p. 456–457)

To summarize thus far, based on an examination of research conducted within behavior analysis, the generality of some widely accepted notions about punishment may be questioned. Unlike many pseudoscientific ideas, these notions are grounded in empirically informed principles. Yet, if Cipani (2004) is correct in asserting that the research evidence regarding unpleasant side effects of punishment has been overinterpreted, then as a community we behavior analysts

have fallen short of our own prescription to “judge the validity of a claim based on objective empirical evidence” (Normand 2008, p. 42). The resolution of this issue awaits new basic and applied research that behavior analysts are perfectly capable of conducting (e.g., Lerman and Vorndran 2002). Unfortunately, with little new research on punishment currently being conducted within behavior analysis, the wait for this new evidence may not be brief.

Evidence About Punishment from Outside of Behavior Analysis

Some fundamental behavioral concepts, including punishment, have made their way out of behavior analysis and into the mainstream of scholarly thought in the behavioral sciences, and so it should come as no surprise that non behavior analytic researchers and theorists have invested considerable effort into understanding the punishment process. For example, several recent large-scale literature reviews have sought, and largely failed to find, the adverse emotional and developmental effects of mild corporal punishment that might be predicted based on the Second Core Truth (Gershoff 2002; Larzelere and Kuhn 2005; Paolucci and Violato 2004). Among the many studies reviewed (e.g., 157 of them in Paolucci and Violato 2004), only a tiny handful appeared in behavior analytic journals, and none were referenced in the best recent behavior analytic review of applied punishment (Lerman and Vorndran 2002). This is concerning because a field that is serious about skepticism should seek out empirical evidence wherever it may be found and be intensely interested in effects that might prompt the revision of established beliefs. The nearly complete segregation of behavior analysis from mainstream psychology, however, means that a sizeable pool of evidence that might inform our theories and principles of punishment is being systematically ignored.⁴

A Possibly Overlooked Function of Punishment

Another way in which “outside” evidence might challenge us is by suggesting punishment effects that behavior analysts

⁴ Some of my colleagues might discount this external evidence due to the fact that relevant studies were conducted using experimental methods in which they have limited faith. While part of skepticism does require considering the quality of empirical evidence, the wholesale dismissal of a literature rarely enhances a discipline’s credibility. I am reminded here of Hearst’s (1967) characterization of the typical basic behavior analyst as “a hard-nosed experimentalist who ... attacks anything that sounds even mildly theoretical or physiological, ridicules anyone who has ever used statistics of the R.A. Fisher variety, and ignores the work of any psychologist who does not publish in the *Journal of the Experimental Analysis of Behavior*” (unpaginated).

have not thought to propose. Findings from the experimental study of cooperation may provide this kind of evidence.

Cooperation may be defined generally as “collaboration... in a collective effort that creates more value than it expends” (Lucas 2006, p. 130). The implication is that when humans cooperate on joint endeavors they can accomplish more good than could be generated by any single individual. Much of what characterizes modern society — government, public works projects, philanthropic organizations, international diplomacy, and the like — is the result of the coordinated efforts of many individuals, and among species humans seem to be unusual in the extent to which they engage in cooperative ventures. Humans regularly band together for the common good, whereas the norm appears to be much lower for many other species (“I should like to see anyone, prophet, king, or God, convince a thousand cats to do the same thing at the same time”; Gaiman 1990, p. 23). Some observers have concluded that cooperation helps to define the essence of humanity (Bowles and Gintis 2013, called us the “cooperative species”), or at least is central to the best of the human world (e.g., Gintis 2008; Russell 1954/2009).

Fragility of Cooperation

The curious thing about human cooperation is that while it seems rife in the everyday world it is surprisingly difficult to capture in the laboratory (e.g., Fehr and Gächter 2002). To explain this point, it will be helpful to briefly describe a procedure that researchers have devised for modeling cooperation in the laboratory, called the public goods game (e.g., Andreoni 1988). There are other cooperation laboratory tasks, including the familiar Prisoner’s Dilemma game (Dawes 1980), and many of the general findings that I will present below for the public goods game have been replicated with other procedures.

Some key mechanics of the typical public goods game are explained in Fig. 1. All laboratory cooperation tasks involve multiple participants, or players, and joint contingencies in which each person’s outcomes depend on the combined actions of all players. Players have the option to compete (act in their own selfish interest) or cooperate (act for the “greatest common good”). Contingencies are arranged so that individual benefits are maximized through competition, whereas the sum total of benefits for all players is maximized when everyone cooperates. Benefits to all are reduced when all compete. As Dawes (1980) has described, such contingencies model the key properties of many everyday social dilemmas:

People asked to keep their thermostats low to conserve energy are being asked to suffer from the cold without appreciably conserving the fuel supply by their individual sacrifices; yet if all keep their thermostats high, all may run out of fuel and freeze.... During pollution

alerts... residents are asked to ride their bicycles or walk rather than to drive their cars. But each person is better off driving, because his or her car’s contribution to the pollution problem is negligible.... Yet all residents are worse off driving their cars and maintaining the pollution they would be if they all bicycled or walked (p. 1970).

While some cooperation is observed in laboratory models, it often occurs unreliably (e.g., Fehr and Gächter 2002), even when the payoff matrices are adjusted to make cooperation more profitable than shown in Fig. 1.⁵ Some individuals do not cooperate, and those that do typically do not cooperate so all the time. Cooperation may be scarce early in a game and develop gradually or, more typically, be evident early in a game and evaporate as the game progresses. Overall, a potential implication is that positive reinforcement contingencies are insufficient to maintain cooperation under conditions where there are individual benefits for competing (Boyd and Richerson 1992) — a rather strange state of affairs for a “cooperative species.”

Punishment’s Effects

A reliable finding, however, is that *adding punishment contingencies* to games like that shown in Fig. 1 promotes reliable cooperation (Fehr and Gächter 2002). One common approach is to allow players, following each trial or round of the game, to assign monetary penalties to other players of whose choices, for whatever reason, they disapprove. Perhaps not surprisingly given the often-noted human proclivity for administering punishment (e.g., Skinner 1971), players take advantage of the punishment contingency when it is available to them, but they do not do this indiscriminately. The majority of penalties follow others’ acts of non-cooperation, suggesting a tendency for people to pay special attention to violations of “fair” social exchange (Fehr and Gächter 2002; Herrmann et al. 2008). An obvious possible implication is that tasks like the public goods game do not fully model the circumstances that produce everyday cooperation. Presumably, where human cooperation is observed in the everyday world, there exist supplemental contingencies beyond those that define a social dilemma — including contingencies that punish non-cooperation.

It may seem obvious that players of cooperation games would punish non-cooperation that adversely affects their earnings. An interesting thing happens, however, when players cannot assign penalties and this option is offered

⁵ Cooperation can be increased by making it the most individually profitable strategy, but under such circumstances individual self-interest aligns with the common good and the task no longer parallels the social dilemmas that cooperation games were intended to model.

Fig. 1 The public goods cooperation game. The game is meant to simulate real-world cases in which investment of effort or resources by many people into a joint endeavor yields enhanced outcomes for everyone. In this example version, four people (*person icons*) each begin the game staked with an amount of money, say \$20, and may keep any money still in their possession at game's end. In each round of the game, a player can choose to keep \$1 for individual purposes, or invest (*thick arrows*) \$1 in a "community project" that yields benefits for all players. Imagine, then, that each player, regardless of whether he or she contributes, receives \$0.40 (*thin arrows*) for each dollar that is contributed in each round. Cooperative behavior is measured in terms of contributions to

the community project. Uncooperative behavior (sometimes called "free riding") occurs when a player does not contribute but reaps project benefits anyway. For an individual, it is more profitable to free ride than to cooperate. For example, if three of four players contribute, each pays \$1 and receives \$1.20 (\$0.20 profit), while the fourth (free riding) player pays nothing but earns \$1.20 (\$1.20 profit). However, the "greatest good for the greatest number" is achieved if all cooperate. With no investors over 20 rounds of the game, four players each keep their original \$20 stake for a total "net worth" of the community of \$80. But if there are four investors in each round, the community's "net worth" is \$128, or \$32 per player

instead to a "bystander" who does not play the game and cannot gain money from it (Alenberg et al. 2011). Although no contingency demands it, the bystander typically attends closely to the game and assigns penalties, again not indiscriminately but instead mostly contingent on acts of non-cooperation. Here, the bystander does not benefit from assigning penalties, but through selective application of penalties may assist other individuals by helping them achieve the "greatest common good."

In games of cooperation, both players and bystanders will assign penalties to non-cooperators even when doing so comes at a personal cost. For example, in some versions of the games, an individual's own earnings are reduced every time a penalty is assigned to someone else. The resulting behavior, called "costly punishment" or "altruistic punishment," occurs nearly as often as when no personal cost is involved (Alenberg et al. 2011). In other words, this tendency appears to be quite resistant to change.

In general, punishment of non-cooperation is a robust phenomenon. Figure 2 shows results from a study by Herrmann et al. (2008) in which adults in many cities, representing many cultures, played a public goods game.

Fig. 2 Money invested in penalizing other players in a public goods game. A player had to deduct from his or her own money to "purchase" the right to penalize another. Expenditures on penalizing are shown as a percentage of the amount to which a player originally was staked. Data replotted from Herrmann et al. (2008), Fig. 1

The amounts of money involved were normed to local economies, and players could penalize others as desired, at a personal financial cost. The black bars show that rates of penalizing non-cooperation were rather similar across societies. The white bars show highly variable rates of what the authors called “antisocial punishment,” which refers to penalizing following any other action. One way of interpreting these findings is that punishing non-cooperation is something of a human universal, whereas punishment used in other ways is a function of locally idiosyncratic cultural practices (though see Heinrich et al. 2006, for possibly contradictory findings).

Because people will make personal sacrifices in order to punish non-cooperation, and because the overall prevalence of this type of punishment appears to be unaffected by local cultural practices, it has been proposed that perhaps humans have evolved to find the administration of punishment for non-cooperation reinforcing (Boyd and Richerson 1992). In support of this hypothesis, imaging studies show activity in the brain’s “reward centers” when players of cooperation games penalize non-cooperation — but not when they administer penalties for other reasons (de Quervain et al. 2004).

To summarize the present section, the cooperation literature suggests the capacity for people to benefit in two ways when they penalize non-cooperation. First, for persons directly involved in collaborative endeavors, the penalties may promote enhanced cooperation by others and thus promote the “greatest common good.” Second, penalizing others’ non-cooperation may even may feel good (be automatically reinforcing). Perhaps not surprisingly, there is evidence that people prefer to be part of collaborative endeavors in which altruistic punishment is possible. Gurek et al. (2006) allowed people to choose between two public goods games. In one, players could administer penalties to each other and in the other no penalties were allowed. The games proceeded for 30 rounds, and after each round a player could choose to switch to the other version of the game or remain in the one played most recently. As Fig. 3 shows, initially a majority of players chose to be part of the no-punishment game, but gradually, across rounds, nearly all players switched to the game that included punishment.⁶

⁶ These results do not indicate whether the punishment game was preferred because penalties enforce cooperation, because it “feels good” to punish non-cooperation, or because of some combination of these factors. The noteworthy point, however, is that almost all players chose to place themselves in a game where they could be penalized. These findings have limited precedent in the behavior analysis literature. In a few cases, individuals have been shown to prefer contingencies involving aversive stimulation to contingencies involving only positive reinforcement (e.g., Hanley et al. 2005; Iwata and Bailey 1974). In these cases, however, the contingencies were strictly individual, and beneficial outcomes were measured at the individual level. The present focus is on preference for punishment-based social contingencies.

Possible Implications

Taken at face value, the findings just described suggest potential challenges to our traditional conceptions of punishment. Most notably, although in behavior analytic writings punishment often is depicted as a destructive force that undermines human society and alienates the individuals who comprise it (e.g., Sidman 2000; Skinner 1971), in cooperation studies punishment appears to serve as a hedge against selfish interests. That is, punishment may be one factor that allows humans to collectively create better outcomes, through cooperation, than they could create individually (Boyd and Richerson 1992), and its beneficial effects may sometimes attract people to environments where punishment operates.⁷

If humans really are a “cooperative species” and humanity’s greatest achievements really arise through collaborative effort, then we are confronted with a very different perspective on punishment than usually seen in behavior analytic sources. Punishment, rather than being one of society’s great ills (Sidman 2000; Skinner 1971), might sometimes function as an essential adhesive that helps to hold society together.⁸

Concluding Observations

The present discussion does not necessarily imply that the Second Core Truth about punishment is wrong. Punishment could serve simultaneously as an engine of collaborative effort, as per the cooperation literature, and a scourge of individual well being, as per the Second Core Truth. Nothing requires the “greatest common good” to be optimal for any individual. My message, however, does not really

⁷ The “better outcomes” are defined socially; that is, “greatest common good” means maximizing collective, not individual, benefits. In both laboratory and clinic, behavior analysts usually seek to identify functional relations governing individual behavior. Thus, the available evidence on not just punishment, but most behavioral phenomena, comes from this level of analysis. One wonders what other behavioral phenomena, besides punishment, might look different to us if viewed through the lens of interactive social relations.

⁸ To those who find this conclusion about the potential social benefits of punishment unsettling, take heart: Research also indicates that when given the opportunity players in cooperation games will reward others’ cooperation, even when this costs them some of their own earnings. Such player-mediated rewards tend to increase the overall level of cooperation in the game (Balliet et al. 2011; Szolnoki and Perc 2010). As with the punishment contingencies I have described, however, these are “additional” contingencies, superimposed upon the primary contingencies embedded in the game, and thus do not change the general observation that cooperation is unreliable under typical payoff systems modeling social dilemmas. Additionally, cooperation is enhanced more by allowing players to penalize non-cooperation than by allowing them to reward cooperation (Baillet et al. 2011). Thus, my general point about the potential value of punishment appears to hold.

Fig. 3 Preference for a cooperation game including versus excluding an option to penalize other players. Data replotted from Gurek et al. (2006), Fig. 1

hinge on whether the Core Truths about punishment are incorrect or incomplete. The critical point is that, given the current state of the evidence, it may be hard to tell for sure. At the least, a consideration of evidence — as skepticism demands — opens the door to an important conversation about our “established truths” regarding punishment.

There is no reason why our professional community cannot respect the strong tradition of research and theory that has taught us much about punishment and simultaneously remain open to the possibility that evidentiary foundations of punishment principles are not fully settled. Indeed, as any elementary textbook on research methods will remind us, respecting evidence means being willing to modify one’s beliefs as evidence may demand.

I have suggested that punishment principles, as behavior analysts typically express them, may be questioned on two grounds — due to possible translational overconfidence regarding the external validity of unpleasant side effects that have been detected in the laboratory, and due to a failure to examine all of the available evidence regarding punishment’s potential impact on the fabric of society. In light of the passions that surround clinical applications of aversive control, my observations are likely to be controversial, and it is reasonable to ask how I would have the reader act on them. My modest proposal is that we position ourselves among “those who accept dogmas and know it” (Chesterton 1923, p. 86). As behavior analysts who cherish scientific traditions, let us, for how, continue to embrace the traditional Core Truths about punishment. Let us teach them to our students (along with the evidence that bears on them) and limit the use of punishment in our personal affairs and clinical work. As fans of the continuing evolution of behavior analysis, however, let

us be open to a possible future in which different Core Truths guide our science, practice, and personal lives. As skeptics, let us work to be certain that research evidence, carefully and thoroughly evaluated, always serves as the arbiter of what we consider to be our Core Truths. In the case of topics like punishment, which have received considerable attention in mainstream behavioral science, this almost certainly means being widely read and keeping abreast of developments from outside of behavior analysis.

References

- Ahearn, W. H. (2010). What every behavior analyst should know about the “MMR cases autism” hypothesis. *Behavior Analysis in Practice*, 3(1), 46–50.
- Adcock, J. (1995). The belief engine. *Skeptical Inquirer*, 19(3). Found online at http://www.csicop.org/si/show/belief_engine/.
- Alenberg, J., Dreber, A., Apicella, C. L., & Rand, D. G. (2011). Third party reward and punishment: group size, efficiency, and public goods. In N. M. Palmetti & J. P. Russo (Eds.), *Psychology of punishment* (pp. 1–19). New York: Nova Science.
- Andreoni, J. (1988). Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics*, 37, 291–304.
- Azrin, N. H., & Holz, W. C. (1966). Punishment. In W. K. Honig (Ed.), *Operant behavior: areas of research and application* (pp. 380–447). Englewood Cliffs: Prentice-Hall.
- Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: a meta-analysis. *Psychological Bulletin*, 137, 594–615.
- Bowles, S., & Gintis, H. (2013). *A cooperative species: human reciprocity and its evolution*. Princeton: Princeton University Press.
- Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizeable groups. *Ethology and Sociobiology*, 13, 171–195.
- Catania, A. C. (2007). *Learning* (interim 4th ed.). Cornwall-on-Hudson: Sloan.
- Chesterton, G. K. (1923). *Fancies and fads*. London: Methuen.
- Chok, J. T., Reed, D. D., Kennedy, A., & Byrd, F. L. (2010). A single-case experimental analysis of the effects of ambient prism lenses for an adolescent with developmental disabilities. *Behavior Analysis in Practice*, 3(2), 42–51.
- Cipani, E. (2004). *Punishment on trial: a resource guide to child discipline*. Reno: Context Press.
- Critchfield, T. S. (2011). Interesting times: practice, science, and professional organizations in behavior analysis. *The Behavior Analyst*, 34, 297–310.
- Critchfield, T. S., Doepke, K. J., & Campbell, R. L. (in press). Origins of clinical innovations: Why practice needs science and how science reaches practice. To appear in F.D. DeGenarro Reed & D.D. Reed (Eds.), *Bridging the gap between science and practice in autism service delivery*. New York: Springer.
- Critchfield, T. S., & Rassmussen, E. R. (2007). It’s aversive to have an incomplete science of behavior. *Mexican Journal of Behavior Analysis*, 33, 1–5.
- Critchfield, T. S., & Reed, D. D. (2004). Conduits of translation in behavior-science bridge research. In J. E. Burgos & E. Ribes (Eds.), *Theory, basic and applied research, and technological applications in behavior science: conceptual and methodological issues* (pp. 45–84). Guadalajara: University of Guadalajara Press.

- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River: Pearson.
- Daniels, A. C., & Daniels, J. E. (2006). *Performance management: changing behavior that drives organizational effectiveness*. Atlanta: Performance Management Publications.
- Dawes, R. (1980). Social dilemmas. *Annual Review of Psychology*, 31, 169–193.
- Deluty, M. Z. (1976). Choice and the rate of punishment in concurrent schedules. *Journal of the Experimental Analysis of Behavior*, 25, 75–80.
- de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, 305, 1254–1258.
- de Villiers, P. A. (1980). Toward a quantitative theory of punishment. *Journal of the Experimental Analysis of Behavior*, 33, 15–25.
- Dinsmoor, J. A. (1954). Punishment I: the avoidance hypothesis. *Psychological Review*, 61, 34–46.
- Dinsmoor, J. A. (1998). Punishment. In W. T. O'Donohue (Ed.), *Learning and behavior therapy* (pp. 188–204). Needham Heights: Allyn & Bacon.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 405, 137–140.
- Gaiman, N. (1990). *Dream of a thousand cats*. Baldwin: DC Comics.
- Gershoff, E. T. (2002). Corporal punishment by parents and associated child behaviors and experiences: a meta-analytic and theoretical review. *Psychological Bulletin*, 128, 539–579.
- Gintis, H. (2008). Punishment and cooperation. *Science*, 319, 1345–1346.
- Gurek, O., Irlenbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science*, 312, 108–111.
- Hanley, G. P., Piazza, C. C., Fisher, W. W., & Maglieri, K. A. (2005). On the effectiveness of and preference for punishment and extinction components of function-based interventions. *Journal of Applied Behavior Analysis*, 38, 51–65.
- Hearst, E. (1967). The behavior of Skinnerians. *PsycCRITIQUES*, 12(8) (electronic).
- Heinrich, J., et al. (2006). Costly punishment across human societies. *Science*, 312, 1767–1770.
- Herrmann, B., Thoni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, 319, 1362–1367.
- Iwata, B. A., & Bailey, J. S. (1974). Reward versus cost token systems: an analysis of the effects on students and teacher. *Journal of Applied Behavior Analysis*, 7, 567–576.
- Krantz, D. L. (1972). Schools and systems: the mutual isolation of operant and non-operant psychology as a case study. *Journal of the History of the Behavioral Sciences*, 8, 86–102.
- Larzelere, R. E., & Kuhn, B. R. (2005). Comparing child outcomes of physical punishment and alternative disciplinary tactics: a meta-analysis. *Clinical Child and Family Psychology Review*, 8, 1–37.
- Lerman, D. C., Sansbury, T., Hovanetz, A., Wolever, E., Garcia, A., O'Brien, E., & Adepape, H. (2008). Using behavior analysis to examine the outcomes of unproven therapies: an evaluation of hyperbaric oxygen therapy for children with autism. *Behavior Analysis in Practice*, 1(2), 50–58.
- Lerman, D. C., & Vorndran, C. M. (2002). On the status of knowledge for using punishment: Implications for treating behavior disorders. *Journal of Applied Behavior Analysis*, 35, 431–464.
- Lindscheid, T. R., Iwata, B. A., Ricketts, R. W., Williams, D. E., & Griffin, J. C. (1990). Clinical evaluation of the Self Injurious Behavior Inhibiting System (SIBIS). *Journal of applied Behavior Analysis*, 23, 53–78.
- Lucas, J. R. (2006). *Broaden the vision*. Westport: Praeger.
- Malott, R. M., Lyon, D., & Malott, M. E. (2002). A history of the association for behavior analysis. *ABA Newsletter*, 23(3), 5–16.
- Metcalfe, J., & Weibe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition*, 15, 238–246. doi:10.3758/bf03197722.
- Mowrer, O. H. (1940). An experimental analogue of “regression”, with incidental observations on “reaction formation”. *Journal of Abnormal and Social Psychology*, 35, 56–87.
- Normand, M. P. (2008). Science, skepticism, and applied behavior analysis. *Behavior Analysis in Practice*, 1(2), 42–49.
- Paolucci, E. O., & Violato, C. (2004). A meta-analysis on the affective, cognitive, and behavioral effects of corporal punishment. *The Journal of Psychology: Interdisciplinary and Applied*, 138, 197–221.
- Pryor, K. (2002). *Don't shoot the dog! The new art of teaching and training*. Reading: Ring Press.
- Rachlin, H., & Herrnstein, R. J. (1969). Hedonism revisited: on the negative law of effect. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 83–109). New York: Appleton-Century-Crofts.
- Russell, B. (1954/2009). *Human society in ethics and politics*. New York: Routledge.
- Rutherford, A. (2009). *Beyond the box: B.F. Skinner's technology of behavior from laboratory to life, 1950s–1970s*. Toronto: University of Toronto Press.
- Sarafino, E. P. (2012). *Applied behavior analysis: principles and procedures for modifying behavior*. New York: Wiley.
- Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.
- Sidman, M. (2000). *Coercion and its fallout* (revised ed.). Boston: Authors Cooperative.
- Skinner, B. F. (1938). *The behavior of organisms: an experimental analysis*. NY: Appleton-Century.
- Skinner, B. F. (1948). *Walden Two*. New York: Macmillan.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1971). *Beyond freedom and dignity*. New York: Knopf.
- Skinner, B. F. (1993). A world of our own. *Behaviorology*, 1, 3–5.
- Szolnoki, A., & Perc, M. (2010). Reward and cooperation in the spatial public goods game. *Europhysics Letters*, 92(3) 38003. Electronic: <http://www.epljournal.org>
- Thorndike, E. L. (1898). *Animal intelligence: an experimental study of the associative processes in animals* (Psychological Review, Monograph Supplements, No. 8). New York: Macmillan.
- Thorndike, E. L. (1913). *Educational psychology* (The fundamentals of learning, Vol. 2). New York: Teachers College Press.
- Thorndike, E. L. (1932). *The fundamentals of learning*. New York: Teachers College Press.
- Ulman, J. D. (1993). The Ulman–Skinner letters. *Behaviorology*, 1, 47–54.
- Van Haaren, F. (2009). “Primum non nocere”: A review of Taking America off Drugs: Why Behavioral Therapy is more effective for treating ADHD, OCD, depression and other psychological problems by Stephen Ray Flora. *Behavior Analysis in Practice*, 2(2), 58–62.
- Vargas, J. S. (2009). *Behavior analysis for effective teaching*. New York: Routledge.
- Wheeler, J. J., & Richey, D. D. (2014). *Behavior management: principles and practices of positive behavioral supports*. Boston: Pearson.
- Wyatt, W. J. (2009). Behavior analysis in the era of medicalization: the state of the science and recommendations for practitioners. *Behavior Analysis in Practice*, 2(2), 49–57.