Chapter 4: Consumer Revolt and the Rising Cost of Defection

ABSTRACT: Consumer defection from direct-marketing tactics like traditional banner advertising has served as a catalyst for more collaborative marketing formats to evolve. Paid search advertising represents an important evolutionary step, because it involves consumers directly in assessing the quality of content and compelling advertisers to cooperate by providing relevant content. The growth in opportunities for direct consumer feedback has produced grim lessons for marketers, as consumers are able to take punitive measures against brands that defect. Instances of consumer backlash through social media, often in the form of viral videos that counter brand messaging, have produced tangible results in compelling marketers to cooperate with consumer demands. Consumers' eagerness to punish defecting brands has both a historical basis – in marketing's long history of defection – and a neurological basis, as individuals take pleasure in enforcing rules of engagement.

As most of the last chapter was spent looking at cooperative strategies through the bird's eye of theory, it's time to come down to earth and put theory into practice in real-life examples. When I speak of an evolutionary theory of marketing, I confess that I am mainly interested in creating a simple counterpoint between traditional direct marketing and the transformations taking place in social media, because my aim is to make game theory useful to social media marketing, rather than to provide a full history of how we got here. But making the leap from direct marketing to social media marketing is a bit like flipping an evolutionary switch from Cro Magnons to modern humans, with no acknowledgment of the steps in between.

So it's worth spending a few pages on how other marketing tools fit into the evolutionary chain of events. I've already noted how banner advertising itself has evolved, primarily by two means: 1) by forsaking (at least partially) the click-through metric that provided a false equilibrium and encouraged defection, and 2) by making itself more relevant through better

E. Anderson, *Social Media Marketing: Game Theory and the Emergence of Collaboration*,
DOI 10.1007/978-3-642-13299-5_4,

targeting. Those changes are prescriptive of changes in digital marketing as a whole, and so they're important to how social media marketing emerged. I'll examine them in order.

4.1 The Unreasonable Standard of Immediate Action

Fundamentally, the click-through metric fails because it saddles advertising with a task that it was never meant to fulfill at all, i.e., to be so persuasive as to divert the consumer from whatever they were doing and get them to do something else, i.e*., to willingly absorb even more marketing messages* on a Web site, and often to make a purchase. Since this behavior runs competely contrary to consumers' attempts to *reduce* exposure to advertising messages, it seems entirely unintuitive.

It also ignores decades of market research on the role of advertising in consumer persuasion. The classic debate is whether advertising is a "strong force," capable of changing consumer brand preference, or a "weak force," capable only of reinforcing or defending brand preference (White 1999). Notably, neither of these two schools of thought envisions a role in which advertising's persuasive force is sufficient to induce immediate action, and that the advertising itself is the conduit for that action; such a role might be termed "overwhelming force." The chasm between what advertising can do and what it was being asked to do in the case of click-through metrics means that the death-spiral described in the previous chapter was inevitable. As long as the cost for cooperation to the consumer – stop what you're doing, respond to this ad – was higher than the potential reward, consumers would continually defect, and marketers would desperately chase their losses through reciprocal defection, i.e., with more pernicious advertising.

By rejecting, at least to a large extent, the "overwhelming force" role for banner advertising, marketers have once again made it a viable medium. Consumers now accept that a certain amount of banner advertising is necessary to enjoy free content, and marketers accept that banner advertising may have other beneficial effects besides inducing immediate action.

I raise this issue because the acknowledgement that immediate action is an unreasonable standard for online marketing was a crucial evolutionary step that allowed social media marketing to evolve. More reasonable stakes for both marketer and consumer open the door for cooperation, because both sides can realize a long-term benefit. (Unfortunately, some practitioners of social media marketing are now calling for immediate-action standards, and if successful, the effects will be ruinous. Those who cannot remember the past… you know the rest. I'll address this dilemma in the final chapter.)

The second evolutionary change in online marketing – enhanced relevance through better targeting – plays an equally crucial role in setting the stage for social media marketing to take hold. As with the withering away of the click-through metric, greater relevance has the effect of lowering the *cost* of cooperation for both sides. Starting from the obvious premise that it is sub-optimal for advertisers to put advertising in front of poor targets, and it is sub-optimal for consumers to be exposed to ads that are poorly targeted to them, both sides gain points in the payoff if the right ads reach the right consumers.

In banner advertising, a plethora of innovations – good aggregated consumer data, behavioral targeting – have made it possible for marketers to achieve an astonishing degree of relevance, so that an ad can be served to the individual consumer based on such granular factors as their past purchase behavior, their content preferences, and their location.

This degree of relevance begins to edge us toward the seemingly impossible 4-4 standard, in which marketer and consumer act in concert in fulfillment of the mutual goal of a satisfying engagement. But advertising itself can never reach this standard, because it always exacts a cost from both parties in creating these moments of potential persuasion. Consumers would always rather have their content commercial-free, and marketers would always rather get consumers for no cost rather than some cost.

But advertising can get *closer* to this standard, and it has done so in an evolutionary step beyond banner advertising and before social media marketing – the missing link, as it were. I refer to the evolution and maturation of paid search as an advertising model.

4.2 The Missing Link in the Evolutionary Chain: Paid Search

Few marketers may recall that paid search was explicitly developed as an alternative to the click dilemma posed by banner advertising. As Daniel C. Fain and Jan O. Pedersen recount in "Sponsored Search: A Brief History," search engine listings were originally monetized with banner advertising in the same way as other free Web content, but this model was especially burdened by the problem of how to get users to click away from relevant content (the search results they were looking for) to less relevant content (the advertiser's pitch).

The solution was to tie the advertising results explicitly to the subject matter for which the user was searching, so that the user might consider the paid content as worthy of a click as the "natural" search results. This further required displaying the paid advertising result alongside the paid results in text links that made them appear more natural. But advertisers were less willing to pay for ad impressions that were merely text links, not graphical ads, and so the "pay per click" (PPC) model evolved, whereby the search engines would only charge the advertisers for actual clicks on the ads (Fain et al 2005).

This model held great advantages for advertisers, since they could extend their relevancy all the way to consumers actively searching for a given subject, and only pay for those consumers whose interests extended to action, i.e., who were ready to click on the ad because its content was relevant. In this way, marketers could cooperate by serving relevant content to interested consumers, and thanks to the PPC model, the reward outweighed the cost of cooperation.

But for the model to work, the reward had to exceed the cost for consumers as well. If some marketers engaged in short-term defection and served up PPC ads that were irrelevant or misleading, e.g., advertising an electric furnace when the consumer searched for a gas one, then consumers would defect in retaliation, learn to ignore PPC advertising, and the entire model would collapse.

But it has not collapsed. In fact, paid search marketing has experienced double-digit percentage growth each year since the adoption of the pay-per-click model in 2003, and is expected to reach nearly $11 billion in 2009 (eMarketer 2009). Some of this growth has come at the expense of banner advertising, with its slower-evolving pay-for-impressions model. Such growth is entirely predictable from a game theory perspective: it offers greater rewards for cooperation than defection, and so it moves both the marketer and consumer closer to the 4-4 quadrant. But the thornier and more relevant question for the evolution of social media marketing is *how* it managed to reinforce rules of cooperation and prevent short-term defections from ruining the system. Understanding how such systems enforce order may have a great deal of relevance to the new systems evolving under social media marketers.

4.3 How Paid Search Gave Free Riders the Boot

Everyone intuitively understands the problems created by spoilers or free riders, because we encounter these problems every day. The person who cuts ahead in the bakery line threatens the whole system, because the people who wait patiently in line may decide that they are suckers for sticking to the rules and defect, causing a run on the pastry counter. The problem is exacerbated in large systems like corporations, where collective rather than individual responsibility makes rule-breaking easier; a competitor who breaks the marketing rules with pernicious tactics and gets short-term results creates a temptation for others to do the same.

In the case of paid search marketing, the search engines themselves act as a governing body, enforcing rules that discourage defection. They have an obvious incentive to do so, since they are fighting to preserve a business model that requires consumer cooperation. To begin with, search engines require relevance and audit their advertisers for compliance, ensuring that the content of the Web site that follows a click on the ad is pertinent to the ad's content. In this way, the rules of participation are made clear to marketers.

Secondly, they offer a system of rewards and penalties for enhanced or degraded relevance. The advertiser's rank – the order in which the ad appears on the page – is determined not merely by the amount the advertiser has bid, but also by something called a "quality score." The search engines assign a quality score based on the popularity of the content with consumers, i.e., by how often a given ad receives a response. Advertisers with higher quality scores achieve higher rankings for less cost, while advertisers with poor quality scores rank lower. Poorly ranking advertisers have a built-in incentive to improve their scores with more relevant content – which includes bidding on only the most relevant keywords – because they'll save money and get more traffic. Consumers are rewarded with more relevant content, and the search engine is rewarded with revenue.

This small feature – search engines' enlistment of consumers in deciding the worthiness of an advertiser – is a groundbreaking yet overlooked development in marketer-consumer cooperation. Historically, advertising systems have been premised entirely on the advertiser's willingness to pay, not on the consumer's interest. Impressions simply went to the highest bidder. Marketers still had a built-in incentive for relevance, since they wouldn't wish to pay to reach the wrong consumers. But such systems still leave wide latitude for defection, as we've seen.

What the search engines uncovered was that a system that engaged marketers and consumers in the mutual pursuit of relevance could benefit all parties. As in all cooperative games, it required a long-term perspective to develop such a system, since a more traditional system based solely on the highest bidder would produce greater short-term revenue and be easier to administer. The new system also worked because it tapped into the broader sociological change brought about by the Web, in which consumers simply demand more control. As we saw in the case of TripAdvisor, the availability of other information sources requires that advertisers compensate for the loss of their informational role with greater relevance and better content.

And for the consumer, the cost of defection has been greatly reduced; if the marketer defects with a low-relevance paid search ad, the consumer can defect in retaliation by choosing a competitive listing, all in a matter of seconds. The search engines merely recognized that consumer engagement was a necessary – if subtle – ingredient in developing an ad system that worked in this new medium, and they made the evolutionary leap forward.

4.4 Enforcing Rules of Cooperation

The search engine model has some important features besides consumer engagement that make it relevant to social media marketing. It is a system with a clear set of rules, rewards for cooperation and punishments for defection, and the involvement of its participants in rule enforcement, so that marketers are less likely to worry about free riders ruining the system while the search engines aren't looking. But it still requires the outside enforcer – the search engine – to mete out rewards and punishments. In the example of the bakery, by contrast, we're apt to see rule-enforcement occur without the intervention of the bakery's staff. A person cutting in line usually *doesn't* cause a run on the pastry counter, because the rule-breaker is called out by others in line and firmly ordered to go to the back, with lots of accompanying dirty looks. Why?

There is, in fact, an entire field of study devoted to answering this question. The study of "common pool resources," pioneered by Elinor Ostrom, draws on game theory but also political and social science to analyze cooperative systems and uncover the common set of best practices for enforcing cooperation. In general, these features include a clear set of rules, rewards and penalties, and a reasonable assurance that defectors can be detected and penalized.

We see all of these features in paid search marketing, but only the third feature – detection of defectors – is in the hands of participants, and only in an indirect way, through the search engines' ability to measure consumer preferences and distribute rewards or penalties in the bidding system. For this reason, paid search cannot be said to be a fully evolved system of marketer-consumer cooperation. The rules of engagement are set by a third party that stands to gain the most by their enforcement. The marketer-consumer interaction is still only a means to an end – a way for consumers to get access to search engines for free, and for marketers to gain an ad platform – rather than an end in itself. In order to evolve the relationship further, marketers and consumers must cooperate in common pool scenarios, where the community of participants enforces its own rules.

4.5 Social Media and the Peculiar Pleasures of Punishment

To show how marketers and consumers are making this last evolutionary step, I need to begin with the concept of punishment. At the risk of being overly reductive, I will contend that marketers' participation in social media's rewards has generally begun with punishment. In consumers' minds, marketers have historically been the line-cutters in the bakery, seeking to maximize their advantage at a cost to the collective good. In the long-term iterative game that is the marketer-consumer relationship, the marketer has gained a reputation for defection, and consumers have uncovered a new form of retaliation, thus changing the stakes of the game. Several well-publicized incidents of punishment have made marketers even warier of social media participation, but such incidents are merely a natural part of the maturation of a mutually rewarding cooperative system, as we'll see.

We've seen already in the case of TripAdvisor that social media venues can provide a powerful reality check on the truth-claims made by marketers, forcing marketers to contend with the power of consumer reviews on their own terms. But social media offers even more powerful means of consumer retaliation in its capacity as an outlet for consumer-generated content.

To understand how this works, first consider how advertising functions at a level of remove from the brand experience itself. Allow me to state the obvious: advertising is not the brand relationship itself, it is a *signal* of what the marketer wishes the relationship to be. Advertising gives marketers the freedom to idealize that relationship in an act of wish-fulfillment; if it is successful (and if one accepts the notion of advertising as a strong force), it prompts the consumer to aspire to such a relationship too. A Nike ad makes the consumer want to be committed and resourceful in pursuit of their fitness goals, and to see Nike products as the means of achieving those goals.

In traditional marketing, marketers have had the brand megaphone mostly to themselves. A brand has the financial resources and the media access to own most of the arguments about what the relationship should be; consumers are relegated to the passive role of accepting or rejecting that argument.

That is, they were until now. A crucial step in the evolution of cooperative marketing is the power consumers now have to dole out punishment to enforce cooperation. Consumers with access to social media can engage in a kind of counter-advertising or counter-signaling, disputing the messages being conveyed about the brand. Doing so raises the stakes of cooperation, forcing the marketer to consider not merely whether and how to advertise but how to improve the customer experience.

4.6 "United Breaks Guitars" Breaks Through

This is best explained by way of example. Let's begin with one of the most prominent incidents of consumer revolt in recent memory, the "United Breaks Guitars" social media phenomenon. In spring 2008, musician Dave Carroll and his bandmates were traveling from Nova Scotia to Nebraska. While the plane was on the tarmac, Carroll witnessed baggage handlers carelessly tossing his band's guitars. Upon landing, he discovered that his prized Taylor acoustic had indeed been broken in transit. In his telling, Carroll began a Kafkaesque series of attempts to gain recompense from the airline. He was repeatedly turned away (United claimed he had not registered his complaint at the correct time for it to be honored), and he vowed to the last United employee to deny his claim that he would exact a musical revenge – three music videos that would expose United's transgressions (Reynolds 2009).

Let's return to the theoretical level for a moment to make sense of what's going on here. In game theory, promises and threats are viable means of reinforcing cooperation: promises indicate a future reward, and threats indicate a future penalty. Threats are particularly useful because the player bears no additional cost if the threat is successful, whereas successful promises must be fulfilled. But for threats to be effective, they must be *credible*, i.e., the recipient must have some reasonable assurance that the threat could be carried out (Dixit & Nalebuff 2008).

Needless to say, individual consumers have not traditionally been able to make credible threats to do damage to a brand's reputation. The consumer would bear great cost in trying to generate sufficient publicity to do harm, with limited chance of success. The economics of customer service account for this low probability; they do not dictate, of course, that the brand does nothing to help unhappy customers, but they also do not add a large multiplier to each unhappy customer, counting on their ability to spread the word of their unhappiness far and wide. According to the *Los Angeles Times*, citing statistics from the Department of Transportation, United Airlines had 13,517 "baggage reports" in April 2009 alone (Reynolds 2009). Like all airlines, United places conditions on its damage compensation policies because the cost of customer irritation is calculated to be less than the cost of making good on every single claim. From a game theory perspective, this is a very rational response, even if it strikes us as cold-hearted or offends our notions of customer service.

Thus in handling customer baggage complaints, United had created, whether they realize it or not, a minimax point, whereby they minimize their maximum loss in paid damage claims, providing compensation in enough cases to keep them out of serious hot water, but not enough to make every customer happy. They were playing a zero-sum game because it was seemingly the best strategy available in this unpleasant corner of the customer experience, which always starts out with unhappy customers. This small zero-sum game had no effect on the larger game being carried out in United’s marketing. Or at least it didn’t until Dave Carroll came along and changed the stakes.

Carroll proved that his threat was more credible than the airline could have possibly envisioned. Possessed of some natural advantages that many online content producers lack – a gift for songwriting, access to good production resources, and a sly sense of humor – Carroll produced a savagely witty song and video with his complaint laid out starkly in the title: United Breaks Guitars. The song quickly became that most elusive of things, a YouTube phenomenon. It logged more than 100,000 views within a few days, and as is typical of such phenomena, it took hold in mainstream media, showing up in countless news accounts, blogs, TV segments, talk shows, etc. As of this writing, the video has garnered 5,656,458 views on YouTube and 22,456 comments – nearly all of them vitriolic in their condemnation of United.

Carroll’s video is only the most prominent example of a broader set of behavior in which ordinary consumers engage in counter-signaling through social media, and they are effectively able to disrupt the signal being sent by the brand through traditional marketing. By escalating his complaint to social media, Carroll transformed his private complaint into an argument about the brand itself and its relationship to its customers. It would be difficult to understate the strength of this signal to social media users; as of this writing, United’s own most recent advertising, by contrast, has garnered 48,315 views on YouTube, or less than 1/100th of Carroll’s audience. Who wins?

The answer to that question is more complicated than it first appears. To anyone who has ever felt deeply wronged by an airline – which is to say, nearly anyone who has ever flown on an airline – Carroll’s musical revenge is deeply satisfying. The fact that Carroll eventually did receive full compensation from the airline is further proof of the rough justice that social media is capable of exacting. But brand-consumer relationships are iterative games, and this is only one round. The “United Breaks Guitars” incident generated a great deal of overheated commentary about consumer revolt, including a claim by Chris Ayres of the *Times Online* that the PR fallout from the incident caused United’s stock to fall 10 percent, “costing shareholders $180 million” (Ayres 2009).

4.7 Recalculating the Cost of Defection

If such incidents did indeed produce $180 million swings in stock values, brands would rightly be in full-blow panic over social media's negative impact, and no equilibrium would even be possible. Consumers could effectively hijack brands with their counter-signals, bringing them to their knees. But that hasn't happened. Analyzing the more subtle effects of the incident is not only more accurate but more useful. Incidents like Carroll's, as well as others we'll examine, first and foremost have the effect of forcing brands to reevaluate how they play the game. United could not use its traditional signaling method – paid media –as a way of correcting the perception created by Carroll's video, i.e., they could not defect in retaliation without making their long-term prospects worse.

Rather, Carroll's original threat was intended to compel cooperation, and it achieved its purpose, so that the long-term prospects of further cooperation between United and its customers are actually improved by the incident. United declared that it intended to use the video in training employees on how to better handle future incidents, and this is perhaps the most important effect of all: it suggests that United now believes such threats are credible and is more likely to make cooperative moves in future iterations, thus bringing its signaling and its actual behavior in closer alignment.

To review: counter-signaling through social media arms the consumer with a new weapon in the iterative marketing game. The consumer's usual means of defection – signal-blocking and brand rejection – are augmented by new punitive tools that raise the stakes of the game and are more likely to compel cooperation in subsequent rounds of play. The counter-signal exacts a cost that is a multiplier on the defection of a single consumer, because it causes other consumers to defect and/or send up their own counter-signals. Marketers' recalculation of the cost of defection becomes an important evolutionary step, one that brings marketers directly into the social media playing field.

But before we consider marketers' forays into social media, there's more to examine on the phenomenon of consumer use of social media to compel cooperation. On a much smaller scale than Carroll's social media juggernaut, consumers have quietly taken up the practice of what amounts to a kind of public shaming of brands that defect from their brand promises. Some of the better-known videos include a cable repairman asleep on a customer's couch, captured on the home's nanny-cam, and a phone recording of an AOL customer service representative that refuses to allow a customer to unsubscribe. By gaining viewership, commentary, and media attention, these counter-signals have functioned as a kind of public shaming of errant players – one of history's oldest forms of punishment.

4.8 Changing the Defection Stakes for Banks: Debtor Videos

It may be stating the obvious to point out the importance of the public or social component to the success of these counter-signals; if Carroll had made a great video that no one saw, United would have been far less likely to act. I raise this point in order to highlight the commodity being contested in these exchanges: social attention is a commodity in finite supply, and consumer revolt works only as long as an adequate supply of this attention can be obtained. But the supply will, in fact, run out, and counter-signaling as a tactic will be diminished. It is a short-term tactic in a long-term game.

How do I know this? I am thinking of the very recent phenomenon of "debtor revolt," in which consumers post videos and write blog entries about their struggles with banks and lenders in the midst of the worst credit crunch since the Great Depression. In the standard narrative, the consumer complains that the lender or credit card issuer won't renegotiate terms that would allow the consumer to continue paying their debt, instead forcing them into default. The more popular videos are the more outrageous cases: a student whose education loan deferment is mistaken for a delinquency, or a landlord seeking a point-and-a-half- reduction in his mortgage rate in order to keep his rentals afloat.

The posting of the video is seen as the logical next step beyond the consumer's failed efforts at 1-to-1 negotiation, but from a game theory perspective, it is much more. As with United Airlines' baggage claim department's handling of Dave Carroll's complaints, these banks and lenders are applying a minimax theorem to the problem of delinquency, setting a threshold at which they believe it will cost them more to extend additional consideration or leniency. On an individual basis this minimax theorem produces outrageous cases, because the lender also loses if an otherwise viable customer defaults over their failure to renegotiate terms. But on a macro level, the minimax theorem prevents the lender from giving up so much ground to consumers that they fall back into the "toxic asset" problem that created the credit crunch to begin with.

For the consumer, then, the debt video is an attempt to force the other player out of the zero-sum game and into one with a cooperative solution by changing the stakes of the game. In this new game, the consumer implicitly argues, the bank's cost of defection is not merely the financial loss – which the bank has already calculated in its algorithms for dealing with delinquency – but the compounded loss of positive brand equity. Without knowing for certain whether the consumer can pay back the loan, the bank believes it has a sub-optimal but stable solution in not renegotiating, and this calculation is correct only insofar as the bank has taken all of its possible losses into account. But the consumer destabilizes that solution by increasing the penalty for defection and compounding the potential loss.

This strategy appears to be working. The landlord received a response from his lender within 4 hours of posting his video, with a promise to investigate his case. Other celebrated cases also received high-level and direct communications from bank personnel empowered to renegotiate. In the case of the landlord, a bank spokesperson claimed that "if he had sent that letter without the video, he would have gotten the same response," but this is unlikely. The bank's calculation changed only when its cost of defection changed, as we would expect (Delaney 2009).

The consumers in these debtor revolt games are winning, at least temporarily, not because their complaints are reasonable (they may be, but that is largely irrelevant), but because they are successfully seizing control of a commodity that the bank brands prize very highly: attention. Positive attention in the form of favorable brand impressions is the commodity the brands purchase in order to win the marketing game, and if consumers can acquire negative attention for that same brand for free, the game changes drastically. But as I noted, attention is a finite commodity in a saturated media environment. A few debtor revolt videos can hold the attention of banks because they can also hold the attention of a sufficient number of Web users and media outlets. But there were 358,471 foreclosure filings in August 2009 alone (RealtyTrac 2009). How scalable is this tactic?

4.9 Attention Saturation and the Limitations of Punishment

Recall that a standard feature of the non-cooperative game is that defection would be disastrous if everyone did it. In the competition for fickle consumer attention, consumer revolt videos on a large scale would quickly lose their luster, which would mean a loss of attention as a commodity. The banks' cost of defection would suddenly be reduced to its previous proportions. Indeed, one can easily imagine consumer backlash against the complainants, because their defection would eventually be perceived as the "free rider" syndrome – players cutting ahead in the bakery line while everyone else has to wait.

I raise the issue of saturation in part because it is, I will contend, the single greatest delimiter of the effectiveness of social media marketing for both marketer and consumer players. Saturation would have a deflationary effect on the value of the video, especially in a medium that places a premium on novelty. I've devoted the final chapter to the issue of saturation and information overload, so I'll explore this issue in greater detail there. The main impact of saturation in this hypothetical case is the sudden drop in the punishment stakes, as brands learn to triage customer complaints that appear in this format.

The optimal solution for consumers as a whole is for this initial defection move – the debtor revolt video – to compel a reevaluation of policies among the banks and a more flexible and accessible approach to individual cases. Since it would be sub-optimal for consumers if the banks either became insolvent or decided they could safely ignore consumer revolt videos, continuous defection would be a very poor strategy. As with Carroll's video, the goal in this game of TIT FOR TAT is to first compel cooperation from the other player, then respond with cooperation of your own.

Unfortunately, consumers don't behave as a monolith any more than marketers do, and coordinating total cooperation would be impossible even if it were a conscious strategy, which it is not. But that is why iteration is so important; as Axelrod showed, the potential for cooperation is directly related to how many times the game is played. As these social media exchanges become more commonplace, brands and consumers both evolve a set of unwritten rules – slowly and painfully, but inexorably – that dictates acceptable behavior, as we'll see in the next example. As debtor videos became more commonplace, *both* players would eventually turn a jaundiced eye on the phenomenon, in the interests of longer-term cooperation.

4.10 The Motrin Moms and Social Media Backlash

My third example of the use of punishment in compelling marketer cooperation is meant to bring us a step closer to a discussion of mutual cooperation, because it involves the increasingly familiar scenario of a marketer earnestly trying to make good use of social media marketing, overstepping the bounds, and incurring consumer backlash. As I noted previously, successful cooperative scenarios invariably require a widely understood, well articulated set of rules – a non-distorted transmission of information between players. Suffice it to say that such rules are neither well understood nor well articulated in the burgeoning social media space, and many brands that take the extra risk of going first take the brunt of the painful lessons for the rest of the industry.

One of the unfortunate side effects of the astonishing viewership of viral videos like Dave Carroll's is that marketers have tried desperately to emulate its success, with widely varying results. The question of what makes a viral video successful is indeed a fit subject for game theory, but it's also complex enough to belong in a later chapter. Here I'll simply offer a brief analysis of motives: marketers that develop their own viral videos can be justly credited with trying to talk to consumers in the social space where the content of their videos can be judged on their own merit, in collaboration with the consumer, as opposed to a video being thrust upon consumers by virtue of the marketer having purchased airtime. In the long continuum of cooperative marketing, it constitutes progress.

But the marketer is also interested in getting something for nothing – namely, ad impressions without media cost. Doing so would obviously improve the marketer's position in the marketing game, because they'd be able to increase the ratio of consumer response vs. the cost of exposure. In paid advertising, the marketer needs a certain number of consumers to like the video (or at least not actively dislike it) in order for the investment to be worthwhile. But in viral advertising, the marketer needs much more: the consumer must genuinely *like* the video in order for it to get any exposure at all, because the marketer relies on the consumer not just as the video's audience but as its distribution channel. This is a crucial lesson: the marketer's stakes are not actually *reduced* by using viral video; the marketer simply *replaces* an investment of actual capital – the cost of running an ad – with social capital, i.e., the effort of engaging consumers through a compelling viral video.

Unfortunately, many marketers that make the leap into viral video have not performed this analysis. They are still playing a mutual defection game focused on maximizing impressions at the lowest possible cost, and they miss the need for collaboration. This is in substance what occurred in one of the most notorious cases of social media backlash, popularly known as the "Motrin Moms" incident.

In November 2008, the pain reliever brand Motrin, a product of McNeil Consumer Healthcare, posted a cheeky viral video to their Web site. The ad poked fun at the trend of moms carrying babies close to their bodies in slings, wraps, or "schwings," as the ad put it. The ad teased that this was an attempt to look like an "official mom" and offered Motrin as a product that could help mothers with the aches and pains of slinging their children.

The tone of the ad was not malicious, and when stepping outside of the controversy, one can even imagine that the ad was intended to "demonstrate genuine sympathy," as McNeil later claimed. But it debuted in a medium that leaves scant room for error in tone, especially where satire is concerned

(Belkin 2008). The ad may well have been unpopular had it run on television, but the consumer's primary weapon of passive non-response – ignoring the ad or changing the channel – would have generated little heat. Online, the ad created not just heat but a wildfire.

Motrin posted the video on their Web site and YouTube on a Friday, and by Saturday, mothers using the microblogging site Twitter began spreading the word among their peer groups about the ad's controversial content. It is the nature of controversies spread through Twitter to grow exponentially, because each recipient of the controversial information can spread the content to a unique circle of friends, who each have their own circle, and so on. Since every consumer equipped with a personal computer is also equipped with the means to make their own video, protest videos began appearing on YouTube by Saturday afternoon. The most popular of them merely showed screenshots of Tweets posted by angry consumers, set to background music. Each of these videos in turn garnered dozens of comments and thousands of views, all in the same weekend. News accounts and blog entries quickly followed, and thousands more consumers with no direct stake in the controversy witnessed the conflagration (Belkin 2009).

The incident is especially chilling for marketers because the controversy reached a fever pitch before Motrin even became aware of it – over the course of a single weekend. There is no precedent for this in traditional advertising; far more controversial ads would take weeks to generate comparable reactions, and the controversy could be expected to ebb as soon as the offending ad stopped airing. In the Motrin case, the ad itself, all protest videos and associated comments, the angry Tweets, and the dozens of blog entries and news articles on the topic are all *still on public display a year later*, accessible by a simple Google search. As a form of counter-signaling, the protest achieved far larger brand awareness and perception effects than the original video could ever hope for. This level of defection is unquestionably a game-changer.

It's easy to see the Motrin incident as a prime example of consumer defection, but the lessons it holds for the marketers depend a great deal on whether one sees Motrin's original viral video as an instance of cooperation or defection. It fits much more neatly into my analysis if we label it "defection," because the logic of TIT FOR TAT then applies: Motrin defected, and Motrin Moms defected in response, and forced Motrin back to cooperation, starting with the lavish apology that appeared on the Motrin home page in the aftermath of the backlash.

Further supporting the "defection" label is the fact that the video itself is so tone-deaf to its audience, despite being reliant on that audience for distribution. Like many marketers, Motrin seems to have misread the signals coming from viral video success, which suggest (distortedly) that "edginess" is a key ingredient to getting a video passed along. In this view, Motrin merely wanted to enjoy the lower hard costs that viral marketing could offer, without being willing to pay the social costs of understanding its audience and crafting something they'd like. And there's a key lesson available in this view: *participation by marketers in social media does not itself constitute an act of cooperation.* It will require deeper engagement, as I'll explore in the next chapter.

But I'm reluctant to consign Motrin to the dustbin of social media defectors so quickly, because the reality is a little more complicated. Social media often behaves like an echo chamber, in which an initial negative reaction pings around endlessly, and subsequent content consumers never see the original content outside of that negative context. In retrospect, Motrin's effort to produce a hip viral video appears hamfisted and slightly embarrassing at worst – like a dad showing off in front of his teenage daughter's friends.

Therefore I propose we assign to the Motrin gaffe the more precise label of "accidental defection," which will in turn describe many of the more notorious social media miscues that have come from brands in recent years. You may recall that accidental defection is the Achilles Heel of the otherwise cooperation-friendly TIT FOR TAT strategy; it can create the endless cycle of retaliation described in the last chapter as the death spiral. Motrin's response to its disastrous first foray into social media could be very long retreat away from the space and back toward its traditional marketing, and this would be a defection, even if it doesn't raise hackles the way the viral ad did. This would be a long-term loss for both players, since Motrin clearly *wanted* to play well in social media, and the "teaching moment" produced by the backlash could help them to do so.

4.11 TIT FOR TWO TATS and the Virtues of Forgiveness

The solution would be for Motrin to play the modified TIT FOR TAT strategy known as TIT FOR TWO TATS, described earlier as a proven antidote to the death spiral. Under this strategy, an extra round of forgiveness prevents a single defection from derailing the entire game. Instead of retreating in response to consumer backlash, Motrin would try again, correcting the mistakes of the previous round.

It could be argued that the burden of TIT FOR TWO TATS belongs with the consumer, since Motrin made the original accidental defection. The consumer should forgive the transgression and give Motrin another chance.

After all, in the case of the embarrassing show-off dad, we do not expect the daughter to disown her father in retaliation. Yet many of the "Motrin Moms" did exactly that, swearing off the product for life. And while their anger certainly seems outsized in retrospect, it's not atypical for acts of consumer backlash in social media. So how does game theory account for such reactions?

4.12 The Enduring Appeal of Punishment

The first explanation is the premise that began this chapter: consumers generally feel that they are the object of a very long-term pattern of defection on the part of brands and marketers, as evidenced by the very low trust in advertising shown in studies like the Edelman Trust Index. Marketers simply cannot treat their entrée into the social media space as though it is a clean slate – a brand new game apart from the one they've played for decades. Consumers will take advantage of the medium's natural suitability for grassroots backlash on the basis of very little provocation. Motrin found this out the hard way.

The second explanation may, in fact, explain a great deal about the behavior of individuals in social networks in general. It springs from the study of player behavior in iterative coordination games, i.e., games in which players must coordinate their moves in order to achieve a common good. Dixit & Nalebuff describe the willingness of players to punish defectors even if the punishment came at their own personal expense. Players given the opportunity to lower their own payoffs in order to impose an even lower payoff on defectors did so *eagerly*. In a purely mathematical view of game theory, such behavior is highly irrational; it is, as Dixit & Nalebuff point out, "a dominated strategy," since the punisher ends up paying more than the group as a whole.

We need to understand this behavior in order to make sense out of much of what goes on in social media. The Motrin Moms who spent the weekend making protest videos ostensibly gained nothing, not even personal fame, from their sacrifice of time and effort. Web community members that troll forums and message boards looking for rule-breakers seem similarly out of balance with their best interests. A player acting purely in their self-interest wouldn't bother with punishment; they would simply tune out or ignore the rule-breaker.

But as it turns out, there is a personal benefit to punishing rule-breakers that balances out the cost, allowing us to reassign this behavior to the realm of the rational. In "The Neural Basis of Altruistic Punishment," published in

Science, a group of researchers from the University of Zurich documented the results of brain scans conducted during cooperative games. They discovered that doling out punishment activated the pleasure centers of the brain in the dorsal striatum, and that even the *anticipation* of being able to punishment defection was pleasurable (de Quervain 1994). Some subjects derived much more pleasure than others, and one can deduce that such subjects are also the people most eager to punish defectors in social networks. (The anticipation factor also explains why some players actively seek out defectors, instead of merely reacting to defection when it crops up.) The bottom line is this: any assessment of the cost of defection and the likelihood of punishment must take into account the existence of these "squeaky wheels" who are eager to punish.

This insight into human behavior would not delight the marketer contemplating a first foray into engagement with recently empowered consumers in the social media arena. If a broken guitar neck or a baby sling joke can ignite a social media firestorm, how does one go about protecting the brand's interests in this new space? At what point do the benefits of participation outweigh the costs of non-participation, and how can that even be gauged?

While I've given a chapter's worth of attention to the concept of punishment, social media marketing will ultimately be defined by its many successes rather than its few failures. It may be cold comfort to United Airlines, Bank of America, and Motrin, but these rounds of defection are as integral to long-term cooperation as the success stories; they advance the creation and transmission of the rules of the game. In the next chapter, I'll explore the next evolutionary step: the success stories that result from the kind of clarity that can only come from failure.