

The Rise and Fall of the 15%: Evolution of Advertising Agency Compensation

Sharon Horsky

Hebrew University, Jerusalem

Robert Zeithammer

University of California, Los Angeles

September 9, 2021

Abstract: We empirically document and analytically model the evolution of advertising agency compensation over the last 200 years. The most enduring compensation scheme—prevailing for over 100 years until the 1980s—was a fixed 15% commission on media billings. This puzzling arrangement has roots in the mid-19th century when agencies worked as media-space salesmen for local newspapers instead of working for advertisers as they do nowadays. Why did agencies switch sides? If they switched from publishers voluntarily, why did they form a collusive association to demand the compensation scheme they just switched away from? How did the commission compensation scheme persist for over a century, and what eventually caused its demise? Our dynamic analytical model explains the switch as arising from the growing potential for national advertising, explains why the switch was irreversible, and characterizes the conditions under which collusively maintaining the pre-switch commission levels benefits the agencies after the switch. The paper's second part then documents that neither government anti-trust rulings nor the proliferation of new media throughout the 20th century were able to break up the collusive arrangement. Instead, its demise followed the unbundling of media buying from creative development in the early 1990s.

Keywords: Advertising, Advertising Agencies, Compensation, Media, Principal-Agent

Contact: Sharon Horsky, sharon.horsky@mail.huji.ac.il

1. Introduction

Our paper models and explains the forces underlying historical shifts in advertising agency compensation. We start with a brief chronological overview of three distinct historical epochs outlined in Table 1. In the first epoch — from the emergence of mass advertising in the 1840s to approximately the 1880s — the agents sold media space on behalf of local newspapers and received a commission on the revenue they generated, consistent with agency theory of salesforce compensation (e.g., Holstrom 1979, Basu et al. 1985). An industry standard emerged, whereby these "sales" agents were paid 15% commission nationwide.

In the second epoch—for about a century, starting in the 1880s —the agencies switched sides to work for the advertisers and put effort into creative work, but they continued to be paid the 15% commission on media billings as in the preceding epoch. At first, observing an agent compensated on a commission basis seems consistent with the large marketing literature on salesforce incentives. But upon further reflection, the second-epoch advertising agency was a buyer of media, not a seller of it like the agents from the salesforce literature. Unlike in the previous epoch, the commission compensation scheme is thus not easily rationalized, because its incentives do not seem aligned with those of the principal (the advertiser): the commission gives neither an incentive to be frugal with the advertising budget nor an incentive to be as creative as possible. We document that this scheme was kept in place by a cartel-like *recognition system* in which the agencies and the publishers colluded to keep advertiser costs high. The collusion survived the emergence of new media such as radio and TV, as well as a 1956 antitrust ruling against it.

Table 1: Timeline of Compensation Plans

Epoch	I: 1840s–1880s	II: 1880s–1980s	III: 1980s–current
Principal	Newspaper Publisher	Advertiser Firm	Advertiser Firm
Agent	Newspaper Agent	Advertising Agency	Advertising Agency
Compensation of Agent	15% commission of media space <u>sold</u>	15% commission of media space <u>bought</u>	Agency’s Cost-Plus (labor-based)

Starting in the late 1980s, the third epoch brought the trend of unbundling of the creative and media services from a single full-service agency to multiple specialized players (Horsky 2006). With the advertisers' media budgets no longer flowing through the agencies working on creative copy, the agencies were no longer able to sustain the collusive arrangement and fairly rapidly acceded to labor-based compensation. Labor-based compensation prevails today, including in much but not all of the digital advertising space (agencies specializing in buying digital media usually charge a commission).

To explain the history outlined above, we first provide a formal analytical model of the first two epochs and then explain the transition from the second to the third epoch using a simple theoretical argument supported empirically by a detailed time series of compensation schemes and agency types.

The first two epochs raise several questions: Why did the agencies switch sides around the time of the emergence of national advertisers? Why—having switched sides—did the agencies then actually prefer to be paid according to their old contracts with the publishers? And why were they unable to get back their old publisher contracts? We now briefly outline the main assumptions and results of our stylized model that answers these questions: our model assumes a nation that consists of two cities, each with local retailers and service providers, and a local publisher who has an audience of local consumers. The retailers and service providers in each city want to let the local consumers know about their goods and persuade them to buy. Because the idea of paid advertising is initially new, potential advertisers require persuasion to buy advertisements in the publishers' newspapers. Publishers do not have the skills to find and persuade potential advertisers, so they contract with independent agents who do. Each city thus initially has a sales agent exclusively working for the local publisher to attract new local advertisers, and the agent is compensated with a share of the advertising revenue as is standard in sales and predicted by contract theory (e.g., Holstrom 1979, Basu et al. 1985).

The key novel element of our model is the accumulation of seasoned advertisers over time. We assume that once a retailer or service provider is “acquired” by the agent to place an ad with the publisher, the publisher easily “retains” the advertiser in the future as a seasoned local advertiser. After all, the publisher merely has to read his own paper to see who the seasoned advertisers are. Although all advertising

is initially local – local advertisers catering to local consumers via a local publisher, we assume that some advertisers have the potential to go national (i.e. advertise in both cities of our stylized nation), and only their local agent knows about this potential and can help them achieve it. As these advertisers with national potential accumulate in the agents’ proverbial rolodexes, the agent in each city has an increasingly lucrative option to quit working exclusively for his publisher and bring some of the advertisers he found to the publisher in the other city (i.e. sell them to the publisher in the other city as new acquisitions). Publishers know about this gradually strengthening outside option of their agents, and so they pay the agents more and more over time until a moment when it is more profitable to just rely on their accumulated seasoned advertisers and let the agents take some advertisers national.

Thus, some advertisers eventually become national in our model at the same time that agents break their contracts with publishers and start representing the national advertisers. The agents have to switch sides and work for the advertisers because they cannot simply go back to working for their publishers at their old lucrative compensation packages: their outside option of taking some advertisers national is used up at that point in time. Because the national advertisers are seasoned in both cities by construction, they do not want to pay the agents for mere representation, and the agents need to provide added value to continue getting compensated by them. We call the added value “creativity” and model it as an increase in advertising effectiveness made possible by the agent’s creative effort. To select an agent, each advertiser holds a contest, in which agents pitch their creative ideas. But how should the winner of the contest be compensated? Does paying their agents a commission on media billings make sense for the advertisers? And how high should the commission be?

At first, advertisers paying their agencies commission on media billings seems wrong, because it seems to give the agent an incentive to spend more on media buying. However, just like a realtor does not decide how much to offer on a house, the advertising agency does not decide the advertiser’s media budget. In other words, the fact that agencies were paid a commission on media billings does not mean the agencies were handed a blank check and a classical salary-plus-commission contract common in sales compensation. In our model, we rationalize the commission structure by assuming the winning agency can capture a

portion of its added value over the runner-up agency in the contest, and this added value is naturally scaled by the size of the advertising budget. We show how competition among agencies erodes the portion of this added value the agency with the best creative idea can appropriate, and we characterize the conditions under which the competitive commission amount is lower than the optimal commission offered by publishers during the previous epochs. We find that once the agents switch sides and start primarily working to provide creative ideas for advertisers, the commission levels naturally arising from the creative pitch contests are often smaller than the commission levels agents used to get from publishers. We thus explain the incentives for the agents to form a cartel with the publishers, reduce entry of new agents (who would drive the competitive commission levels even further down), and successfully demand compensation in line with their old commission levels from when they were working for publishers to acquire new advertisers.

Our formal analytical model covers the first two epochs in the history of advertising agency compensation (see Table 1) and stops short of modeling how the collusive arrangement eventually fell apart in the late 20th century. To model that historical epoch, we rely on Horsky's (2006) explanation of the unbundling of the media buying and creative services, and we document empirically that this infusion of new entrants into the agency market eventually broke the collusive arrangement that characterized the second epoch. To estimate the types of compensation schemes used during the transition from the second epoch to the third epoch, we rely on triennial surveys by Association of National Advertisers (ANA). The time series shows that after the sharp rise in unbundling between 1991 and 1994, the proportion of advertisers using billings-based compensation fell from 61% in 1994 to 10% in 2003, with most advertisers switching to fee-based compensation during the same period.

In this paper, we contribute to the academically under-researched area concerning the relationship between firms and their advertising agencies. The quantitative marketing literature in this area is sparse and includes the following: Gross (1972) and Horsky et al. (2016), who address the design of the competition to select an ad agency; Villas-Boas (1994), who examines the potential benefits had agencies been allowed to service competitive accounts; Silk and Berndt (1993), who find scale and scope economies in advertising agency costs; Horsky (2006), who examines the conditions under which an advertiser should

unbundle the creative and media services to two separate providers; Horsky et al. (2012) who analyze internalization of some tasks, and Wernerfelt et al. (2021) who use the internalization of some task to test different theories of the firm. All of this previous work takes the type of the compensation scheme as exogenously given, whereas we focus here on the ways it has changed over the last centuries.

We are aware of only two papers that address the choice of compensation scheme: Spake et al (1999) who advocate for basing the agency compensation on its added value to advertiser profits instead of using the commission on media billings, and Hao (2005) who focuses on the design of an incentive-based scheme in a standard principal-agent framework modified with an exogenous budget constraint of the principal. These papers are complementary to our as they focus on specific issues in agency compensation while we analyze the overall evolution of it over time.

Within the large field of contract theory (e.g., Jensen and Meckling 1976; Holmstrom 1979; Grossman and Hart 1983; Holmstrom and Milgrom 1987, 1991), our model is most related to Fudenberg and Rayo (2019), who also analyze a dynamic model in which the principal is effectively paying the agent today to develop the agent's outside option for tomorrow. Their principal has an incentive to weaken incentives early so that the agent stays around longer. Unlike their principal, our principal has no commitment and faces a symmetric competitor. To make our analysis tractable, we also assume the principal is myopic, effectively abstracting from the key tension Fudenberg and Rayo study. Instead, we show how the competition gradually unravels the exclusive relationship and drives the agent to look for other employers.

Our paper also contributes to the literature on cartel stability in that we document a stable collusive arrangement that lasted for over 100 years and survived multiple anti-trust rulings. The longevity of the arrangement is surprising given the much more short-lived cartels that started at the same time, such as the Joint Executive Committee (JEC), analyzed by Ellison (1994). We can only speculate about the sources of this stability, and we do so in the Discussion section. The Discussion section also highlight several parallels between the recognition system of the advertising agencies we study and another long-lived collusive arrangement: the National Association of Realtors analyzed by Levitt, Syverson, and Ferreira (2008).

2. Model of the Early History of Agency Compensation

Instead of first presenting our modeling assumptions, then solving the model, and then discussing the relationship between our solution and reality, we present our model chronologically, interspersed with historical vignettes that both motivate our assumptions and closely match our solutions. Although unusual, our presentation strategy is better at parsimoniously highlighting the evolving relationship between the model predictions and historical events, and is also better at introducing additional assumptions and technologies only when they are needed for continuing the evolution of the advertising market. Our historic review of the industry in this and the next section is based on a number of books and articles that examine this time frame: Fowler (1900), Sheldon (1925), Gundlach (1931), Young (1933), Goode (1933), Haase et al. (1934), Klaw (1956), Gamble (1959), Holland (1981), Pope (1983), and the *New York Times* (1926,1927,1956,1958, 1960).¹

History: Agents emerge in 1840s as salesmen of space in newspapers

With the spread of mass-circulation newspapers² in the early 19th century, their publishers understood they could augment their subscription-based income by the sale of print space to potential advertisers. Potential advertisers had to be persuaded to spend money on advertising, and they lacked information about the rates, availability, and circulation of newspapers. Yet, the publishers of the day did not have the skills or bandwidth to solicit paid advertisements: “Newspaper publishers were, for the most part, editors, not business men. They depended primarily on subscriptions for income. Even as important as the Boston Post had no advertising solicitor as late as 1858” (Young 1933, p. 21).

To recruit advertisers, the publishers contracted the sale of print space for advertisements to outside agents, starting in the 1840s.³ These agents were independent middlemen coordinating the market for

¹ A short summary of the main events in the advertising industry history is provided at <http://adage.com/article/special-report-the-advertising-century/ad-age-advertising-century-timeline/143661/>

² Today, except for the weekly *The Economist*, the “newspaper” evokes a daily publication. In the 19th century, most (74%) “newspapers” were published weekly, and they were the “favorite advertising mediums” (Rowell 1872, p. 6).

³ The first known advertising agent was Volney B. Palmer in 1841 in Philadelphia (*Oswega Palladium*, 1846). Advertisements other than local announcements were uncommon in newspapers previous to that time (Young 1933).

newspaper space. Newspaper publishers wanting to generate revenues from selling print space contracted these middlemen to contact potential buyers and compensated them at a fraction of their efforts' output, by using a percentage commission on their sales of space. Thus, the agents' incentives were aligned with those of the publishers because they were compensated based on the amount of media space they sold. In terms of economic theory, the newspaper publishers converged to a compensation plan for their agents, consistent with agency theory (e.g., Jensen and Meckling 1976; Holmstrom 1979; Grossman and Hart 1983; Holmstrom and Milgrom 1987, 1991) and the one advocated in marketing for salesperson compensation starting with Basu et al. (1985). Having outlined the emergence of agents on the advertising scene, we now turn to the model assumptions designed to capture the essential elements of the historical reality.

Model Assumptions: Time

We now begin the description of our model. Table A2 in the Appendix contains the notation for easy reference. Time t starts in the early 19th century and flows discretely and slowly, for example, in decades or other epochs. Because of the long calendar time between periods, we assume all actors are acting myopically, maximizing their payoffs within each period.

Model Assumptions: Lead generation and conversion technology

Suppose an agent exists with the skills required to find and persuade businesses to become newspaper advertisers. When a newspaper publisher contracts with such an agent, the agent works exclusively for the publisher and gets paid both a salary s and a commission c on the new revenue he brings in. The agent's technology for finding and persuading businesses to become advertisers follows the standard set of assumptions in principal-agent modeling to ensure tractability:

- the agent is risk averse with constant absolute risk aversion (CARA) of r ;
- the agent can exert effort e at a quadratic cost $ke^2/2$; and
- the new advertising revenue effort e generates is $\beta(e + \theta)$, where the $\beta > 0$ parameter translates effort to dollars, and θ is a Normal mean-zero shock with variance σ^2 .

The next lemma summarizes the well-known optimal contract implied by the above assumptions:

Model Solution: Application of standard principal-agent machinery to our setting

Lemma 1 (proof in appendix): When \underline{w} is the best change in wealth available to the agent outside of this contract and α is the advertising revenue available to the publisher without the agent's effort, the publisher

sets the commission to $c = \frac{1}{1+kV}$, inducing an effort $\frac{\beta}{k(1+kV)}$, and the publisher makes

$\alpha - \underline{w} + \frac{\beta^2}{2k(1+kV)}$, where $V = r\sigma^2$. The agent makes \underline{w} .

Model Assumptions: Weakly diminishing marginal effectiveness of effort

A natural assumption is that the productivity of an agent's effort in generating new advertising revenue is a non-increasing function of time $\beta_t \leq \beta_{t-1}$ to capture non-increasing marginal returns in finding additional

new advertisers over time. We make a technical assumption that $\beta_t^2 / \sum_{j=1}^{t-1} \beta_j \rightarrow 0$ as $t \rightarrow \infty$, which is easily

satisfied by a range of simplifying assumptions, such as a constant $\beta_t = \beta_{t-1}$ or $\beta_t > \varepsilon > 0$, that is, effectiveness bounded away from zero.

History: National advertising follows the emergence of consumer brands in the late 19th century

In the second half of the 19th century, the advertising agents (also referred to as “jobbers”) enabled advertisers to conduct nationwide promotional campaigns with the same ease with which a local campaign could be conducted. A manufacturer could submit a single advertisement to the intermediary, who then arranged to have it printed in as many different newspapers as was appropriate. Such a service was, in particular, valuable for national brands, which were starting to appear on the national scene (e.g., Heinz Ketchup in 1876, Ivory soap in 1879, or Sears in 1886 ; see Bronnenberg, Dhar, and Dube [2007] for an analysis of the emergence of national brands).

Model Assumptions: Players and geography

A nation exists consisting of two cities $i=1,2$, each with one risk-neutral publisher of the local newspaper. The cities are symmetric in every respect. A continuum of consumer businesses—potential advertisers—

operate in each city, but advertising is a new idea, and they need to be persuaded to pay for it. Not everyone can persuade businesses to become advertisers, and each city has just one agent with such a skill. The agent possesses the lead generation and conversion technology outlined above; the publisher offers the agent a standard incentive contract with a fixed salary and a commission on new advertising revenue generated; and Lemma 1 thus describes the effort the agent puts in, as well as the payoffs of all parties. All advertisers start out as local, meaning they advertise only in their own city's newspaper. But as we discuss next, some of them have the potential to go national.

Model Assumptions: Advertiser life cycle: seasoned and national advertisers

Once a business is persuaded to advertise with the publisher in period t , it is retained forever by the publisher as a “seasoned” advertiser, starting in period $t+1$. For simplicity, we assume the random shock to revenue θ is transitory, and it is the expected revenue from the agent's effort $\beta_t e_t$ that survives to the future as an addition to the installed base of seasoned advertisers. The agent no longer needs to exert effort to bring seasoned advertisers back for another period, and he does not receive commission on that part of the publisher's advertising revenue. In every period, the publisher can thus either hire the agent to generate new advertising business, or rely on his installed base of advertisers for revenue.

All advertisers start out as local, but some proportion v of seasoned advertisers is capable of “going national,” that is, advertising in both cities. The remaining proportion $1-v$ remains local forever. The proportion v is common knowledge, but only the agent in its own city knows which of the local advertisers have the potential of going national; the advertisers themselves are not sure of their own potential.

Advertisers with national potential need agents to actually go national for the first time, both because of the logistics of doing it without precedent, and also because only the agent knows their true potential and can thus assure them about the wisdom of their investment. At any time, an agent from one city can “sell his entire rolodex”⁴ of national advertisers in his city to the publisher in the other city and

⁴ Rolodex was only invented in 1956, 100 years after the period we are discussing here. We nevertheless use the metaphor of the rolodex as a quintessential salesperson's address book with an historical flair.

collect the same commission the distant publisher pays his own agent for finding new advertisers (the commission will be identical in both cities in any period given the symmetry assumption). However, an agent cannot be employed by multiple publishers, so an agent selling his rolodex in the other city foregoes pay from his current publisher that period and needs to renegotiate the contract in the subsequent periods should he want to start working for him again.

Model Solution: Newspaper advertising by local advertisers is born

In any period t , the market is characterized by state (α_t, β_t) as follows: the parameter α_t captures the publisher’s revenue from the installed base of seasoned advertisers, and hence starts with $\alpha_1 = 0$, and evolves according to $\alpha_t = \alpha_{t-1} + \beta_{t-1}e_{t-1}$ before any advertisers go national, $\alpha_t = \alpha_{t-1}(1 + \nu)$ right after all advertisers with national potential go national, and $\alpha_t = \alpha_{t-1}$ afterwards, as long as the agents switch sides and thus no more new business is generated.⁵

In the first period, $(\alpha_1 = 0, \beta_1)$. The agent makes the effort $\frac{\beta_1}{k(1+kV)}$ per Lemma 1. No seasoned advertisers are present and hence no advertisers with national potential; thus, the agent’s outside payoff is $\underline{w}_1 = 0$, where 0 is a normalization meaning “increase in wealth outside of working for the publisher.”

Therefore, the publisher makes $\alpha_1 - \underline{w}_1 + \frac{\beta^2}{2k(1+kV)} = \frac{\beta^2}{2k(1+kV)}$ when hiring the agent, which is more than the zero he would make by relying on the installed base.

Definition: Denote the “channel” profit from newly acquired advertisers net of risk and the cost of effort by $\Pi(\beta) = \frac{\beta^2}{2k(1+kV)}$.

History: From Havoc to the Standardized Open Contract

Although the agents significantly increased the volume of trade that publishers and advertisers transacted

⁵ Adding new “rookie” sales agents in the second epoch to replace the agents who switched sides would be more realistic but would also add clutter to the model without generating any new insights.

with each other,⁶ they were eventually held in contempt by both publishers and advertisers. The problem that arose was the secretive nature of the business conducted between the parties. First, the volume of a publication's circulation determined its advertising rate and publishers tended to overstate their circulation. Second, the agents were secretive about the commission rates they earned from the publishers (which varied between 10% and 50%). The agents were supposed to provide advertisers with impartial advice on which publications best suited their advertising needs. However, advertisers suspected agents were directing them to place ads in publications that paid the highest commissions, rather than those that represented the best advertising value. From the advertisers' perspective, price-gouging and kickbacks seemed to be driving the market for advertising space. We now discuss how the market solved both problems.

The first problem (circulation information) was solved by "Rowell's American Newspaper Directory." Starting in 1869, the directory published circulation figures of over 5,000 newspapers annually, taking care to distinguish verified numbers (described in the directory as "where a willingness has been expressed to make an oath to circulation") from numbers merely "claimed" by the publishers. True to form, Rowell offered paid advertising space after each listing to the paper listed, and many large papers took up the offer. Common advertisements claimed circulation to be increasing and/or largest in a certain geographical area. They also signaled serious journalism with claims such as "no humbug advertisements or quack nostrum notices published" (*De Kalb County News*, 1872) or "inserts no private disease remedy, female pill, or lottery advertisements, at any price" (*Attica Ledger*, 1872). In addition to text right after their listing, many publishers took out larger advertisements to tout the quality of their newspaper. In the early editions of the directory, such advertisements also provided the list prices for advertising space, measured in "squares," equivalent to about 10 lines of text. Interestingly, such transparent list prices slowly vanished from the advertisements during the 1880s, replaced with claims such as "advertising rates furnished on application" or "our rates are reasonable" (Rowell 1884).

⁶ In 1879 56% of newspaper revenue came from circulation sales. By 1919 this percentage decreased to 34%, and the remainder two thirds of newspaper revenues was generated from advertising (Pope 1983).

The second problem (varying commission rates and secretive kickbacks) were addressed by the introduction of the Ayer's "open contract" (OC) in 1875.⁷ The OC guaranteed that the agent would no longer accept rebates or kickbacks from publishers for placing ads with them. Instead, it would accept payment only in the form of a set rate of commission of the list price, a rate that was openly agreed upon between the publisher, agent, and advertiser. This rate varied, experimentally in the beginning, but eventually in 1891 stabilized at 15%, and the OC became standard practice in the advertising industry.

Model Assumptions: No information asymmetry and standard contract

We do not explicitly model the “havoc” period of the industry and instead rely on the OC and Ayer’s directory to motivate our assumptions that the commission rates in both cities are common knowledge in our model, and kickbacks are not allowed.

Model Solution: Agents’ outside option strengthens over time until their publishers no longer want to retain their exclusive services and allow some advertisers to go national

Because of the symmetry between the cities, the commission rates will be identical nationally in our model within every period. We can thus focus on a single city as we investigate the evolution of the advertising market. Because the β_i parameter scales the effort and the shock equally, the commission rates will be stable over time in our model (see Lemma 1), simplifying the analysis and consistent with the historically stable rates of the OC era.

In the second period, the state involves an installed base of seasoned advertisers:

$\left(\alpha_2 = \frac{\beta_1}{k(1+kV)}, \beta_2 \right)$, so the agent now has the opportunity to sell $\nu\alpha_2 > 0$ national advertisers in the other city and make $\underline{w}_2 = \nu c \frac{\beta_1}{k(1+kV)}$. Therefore, the publisher makes $\alpha_2 - \underline{w}_2 + \Pi(\beta_2)$ when continuing to employ the agent. Note how the agent’s outside option of “selling his rolodex” in the other city forces

⁷ By Francis Wayland Ayer of N.W. Ayer & Son advertising agency.

the publisher to give her a higher salary. Relying on the installed base instead would yield the publisher α_2 , which exceeds his profits from hiring the agent whenever

$$\Pi(\beta_2) < \underline{w}_2 \Leftrightarrow \underbrace{\frac{\beta_2^2}{2k(1+kV)}}_{\Pi(\beta_2)} < \nu \underbrace{\left(\frac{1}{1+kV}\right)}_c \underbrace{\left(\frac{\beta_1}{k(1+kV)}\right)}_{\alpha_2} \Leftrightarrow \nu > \frac{\beta_2^2(1+kV)}{2\beta_1}. \quad (1)$$

Note the right-hand side (RHS) of the last inequality drops when the effectiveness of the agent falls over time, as well as when the overall risk parameter V is lower.

Suppose $\nu < \frac{\beta_2^2(1+kV)}{2\beta_1}$ and consider the third period. The installed base increases to

$\alpha_3 = \frac{\beta_1 + \beta_2}{k(1+kV)}$, so $\underline{w}_3 = \nu c \frac{\beta_1 + \beta_2}{k(1+kV)}$. Note how the agent's outside option grows along with the installed

base, but his per-period acquisition of new revenue weakly falls along with his effectiveness. Relying on the installed base in the third period would yield the publisher α_3 , which exceeds the profits from hiring

the agent whenever $\Pi(\beta_3) < \underline{w}_3 \Leftrightarrow \nu > \frac{\beta_3^2(1+kV)}{2(\beta_1 + \beta_2)}$.

Inducting to t period, the publisher who hired the agent in the previous period fires the agent in the

t period as long as $\nu > \frac{\beta_t^2(1+kV)}{2\sum_{j=1}^{t-1}\beta_j}$. Because the RHS approaches zero by our technical assumption, for any

$\nu > 0$, a time will come when the publisher can no longer satisfy the agent's IR constraint and effectively fires her. The agent then exercises her outside option and takes some advertisers national. Given our symmetry assumption, this exercise of the outside option will happen in all cities in the same period.

Period T : advertisers go national.: Let T be the lowest t such that $\nu > \frac{\beta_t^2(1+kV)}{2\sum_{j=1}^{t-1}\beta_j}$. At time T , the

agent's individual rationality (IR) constraint is no longer satisfied, so the agents each prefer to sell their accumulated national advertisers in the other city, and they make \underline{w}_T in doing so. The comparative statics

of T are obvious from the inequality: T increases when fewer advertisers have national potential, when the overall risk parameter V is lower, and when the effectiveness of new-advertiser acquisition diminishes faster with time.

Because $w_T > \Pi(\beta_T)$ by definition, the advertisers make more profit from accumulated national advertisers than the *channel* profit they could generate by staying in their city and finding more advertisers. In other words, taking her rolodex national is not only better for the agent than working for the publisher, but also better than starting her own local publishing business for free.

History: Creativity becomes both feasible for newspapers and important for advertisers

Approximately until the 1880s, agents functioned merely as space brokers; the client supplied the copy and layout, and the agent "placed" the ad and collected payment for the newspapers. The copy and layout were a bold presentation of facts (i.e., mention of available goods) in written form. Many smaller newspapers may not have even had the technology for printing images (called "cuts" at the time) in the advertising columns and qualified their listings in Rowell's directory with "no cuts admitted in its advertising columns" (e.g., *Springfield Times* or *Rhinebeck Gazette* in 1872). Other publications allowed images but charged a premium (e.g., 50% extra in addition to occupied space in the *Christian Advocate*, or "double price for cuts" in the *Saturday Star Journal*, 1872).

In the 1880s, large advertisers started to call on the agents for creative help with the advertising copy, including images and logos⁸: "It was a nuisance to the agent..., but if 'copy' would help sell space then copy he would produce. And if a package, a name, and a trade-mark were also needed to get a new advertiser started, then these, too, the agent would assist with or suggest" (Young 1933, p. 29). At that point in time began the creation of the full-service advertising agency as we know it today—an outfit that performs both creative production and media buying. Gradually, with advertisers increasing demand for

⁸ Large department stores, in particular, consulted with agents for creative help. As the stores' range of goods expanded, and with the increasing ease of consumer travel via new urban streetcar lines, they wanted customers to constantly know what was available and to induce them to come and shop in their stores.

persuasive type advertising, the agencies' emphasis shifted to ad creation instead of media selection: "Ninety per cent of the thought, energy and cost of running our agency goes into copy" (Albert Lasker, ad agency partner in *Lord and Thomas*, 1906). Along with this development, agencies moved from representing the media to representing the advertiser—from selling space to selling clients.

So far, we have shown analytically why the agents left the publishers around the time that national advertisers emerged. To analyze why the agents switched sides instead of returning to work for the publishers, we now enrich our model with the idea of creativity and analyze compensation for the agents it implies.

Model Assumptions: Creative technology

Agents can provide their advertisers value-added services, which we jointly call "creativity." Creativity enhances the effectiveness (i.e., profit to the advertiser per dollar) of advertising spending by some multiplicative factor $1 + \pi$ with the "creativity lift" $\pi > 0$, which is random at the time of investment into creativity, because output of creative effort is unpredictable. We assume an investment in creativity, such as hiring a dedicated artist or testing different versions of copy, thus increases only the distribution of π in the sense of first-order stochastic dominance. For tractability, let $\pi \sim \text{Uniform}[0, M]$, where $E(\pi) = \frac{M}{2}$ is the expected creativity achieved with maximum creativity M . Suppose the agent captures all of π as her payoff, and note a larger M also increases the variance of the creative lift $\text{Var}(\pi) = \frac{M^2}{12}$, which negatively affects our risk-averse agents. To keep the agent's objective function tractable, we assume the cost of achieving a given level of maximum creativity is proportional to the variance of the agent's payoff with a cost-shifter parameter γ scaling the variance analogously to the agent having a higher risk-aversion coefficient. In other words, investing in a higher M feels the same to the agent as becoming more risk averse. When the agent captures all of π as her payoff, the joint cost of bearing the risk and producing the maximum

creativity M is thus approximated by $(r + \gamma) \frac{Var(\pi)}{2}$ in our model.⁹ Given these assumptions, the socially

optimal level of maximum creativity thus solves $M^*(\gamma) = \arg \max_M E(\pi) - (r + \gamma) \frac{Var(\pi)}{2} = \frac{6}{\gamma + r}$.

When the agent works for the publisher, he has no direct incentive to increase the advertiser's advertising effectiveness other than to acquire the advertiser, as explained by the quote from Young (1933) above. We thus make no prediction about the creative effort publishers' agents exert during the first epoch, beyond noting some of the overall effort e corresponds to their creative effort needed to acquire new advertisers for their publisher.

When the agents work for the (national) advertisers, the advertisers hold a creativity contest to choose their agents, and determine the "creativity commission" the agents receive per dollar of the advertiser's advertising spend, presumably something less than the winning agent's π . The exact amount of creativity commission depends on the rules of the creativity contest. The expected creativity commission needs to increase in M to incentivize agents' investment in creativity. Before specifying particular rules of the contest, analyzing the general implications of the creativity commission on the agents' choice of employer in period $T+1$ is illuminating.

Model Solution: When do advertisers switch to working for the national advertisers?

Consider period $T+1$ right after the advertisers with national potential go national. Each publisher now has an installed base of $\alpha_{T+1} = (1 + \nu)\alpha_T$. The agents can no longer get paid a commission on the advertisers they found in the past by taking them to the other city. In other words, with the "rolodex sold," the agent's outside option consists of either zero (a normalization set by $\underline{w}_1 = 0$) or of serving the national advertisers.

Consider one of the agents and let δ be her ex-ante expected creativity commission when competing to work for one of the national advertisers in period $T+1$ *net* of all risk and costs:

⁹ The uniform distribution is different from a normal, so subtracting r times half the variance from the expectation only approximates the impact of the increased risk on the certainty equivalent under CARA.

$\delta = \Pr(\text{win contest})E(\text{commission}|\text{win contest}) - \text{costs}(\gamma, r)$. Note we are not yet making any assumptions about the behavior of the other agent—we do so later when we analyze the creativity contest in more detail. As long as $\delta > 0$, serving the national advertisers is clearly the better of the two outside options for the focal agent.

When will the focal agent’s publisher choose to *not* hire her back? Because $2\nu\alpha_T$ national advertisers exist, the agent’s payoff from working for the national advertisers is $2\delta\nu\alpha_T$. To attract the agent back to finding new advertisers for him, the publisher would thus have to compensate her at least $w_{T+1} = 2\delta\nu\alpha_T$, and himself earn $(1+\nu)\alpha_T - 2\delta\nu\alpha_T + \Pi(\beta_{T+1})$. Not hiring the agent back and simply continuing to serve his installed base of seasoned advertisers yields the publisher $(1+\nu)\alpha_T$ and is thus more profitable whenever $2\delta\nu\alpha_T > \Pi(\beta_{T+1})$. Because $\nu\alpha_T$ is the size of the potential national advertising revenue from one city at the moment of that city’s advertisers going national, it satisfies $c\nu\alpha_T \cong \Pi(\beta_{T-1})$. Therefore, leaving the agent to pursue her creative work for the national advertisers is profitable for the publisher when $\frac{\delta}{c} > \frac{\Pi(\beta_{T+1})}{2\Pi(\beta_{T-1})}$. Because the marginal effectiveness of new-advertiser acquisition is weakly decreasing ($\beta_{T+1} \leq \beta_{T-1}$), the RHS of this inequality is less than half, so we get the following observation:

Observation 1: A sufficient condition for the publisher not re-hiring his old agent in period $T + 1$ is that the agent’s net ex-ante commission from the creativity contest is more than half the old commission for working on acquisition.

When entering the creative market pays, on average, at least half the old commission for acquiring local advertisers, the agents end up working for national advertisers instead of returning to work for their previous employers—the local publishers. Why “half”? The intuition is simple: after advertisers go national, agents working for them can tap into the entire national advertiser market, which is double the

size of each local market in our model. We now specify the details of the creativity-contest game to derive the creative commission δ in terms of model fundamentals.

Model Assumptions: Creative competition, the pitch, and competitive pricing

We model the contest as follows: each national advertiser invites each agency to pitch its creative idea, selects the better one (higher π), and compensates it by its added value over the nearest competitor, namely, pays it the difference between its creativity and the creativity of the losing agency. For example, if agency 1 has creativity lift 0.20 and agency 2 has creativity lift 0.05, the advertiser selects agency 1 and rewards it 15 cents per dollar of its advertising budget. This compensation rule can be micro founded by a Vickrey auction in creativity with the intuitive properties of expected compensation increasing in creativity, and with increasing competition (an increased number of agencies in the contest) decreasing the compensation the winner receives. We propose these intuitive properties are an important feature of real-world pitch contests because the properties give agencies incentives to invest in creativity. However, we do not claim our compensation rule is somehow optimal for the advertisers. Instead, we adopt it as a reasonable and tractable approximation to the payoffs real-world contests are likely to produce.¹⁰

Model Solution: Creative competition and the agents’ switch to working for national advertisers

The agents set their maximum creativity before each contest in a simultaneous-move creative-investment game. Lemma 2 describes the payoff-relevant equilibrium outcomes of this game.

Lemma 2: In the unique symmetric equilibrium of the creativity-contest game, both agents set a maximum

creativity level $M^{**}(\gamma) = \frac{24}{5(\gamma+r)} < M^*(\gamma)$ and collect expected creativity commission $\delta = \frac{16}{25(\gamma+r)}$

net of all risk and costs.

Because the winning agency collects only a portion of its creative lift and wins only half the time, each competitor sets their maximum creativity level below the socially optimal level M^* . Both increased risk

¹⁰ See Horsky, Horsky, and Zeithammer (2016) for more detail on advertising agency contests.

aversion and the increased marginal cost of creativity reduce the equilibrium investment in creativity, and thereby also reduce the expected creativity commission.

When will the publisher choose to not hire his old agent back given our model of the creativity contest? The key $2\delta v\alpha_T > \Pi(\beta_{T+1})$ inequality derived above simplifies as follows:

Lemma 3: In period $T+1$, that is, one period after advertisers go national, both agents end up working for

the national advertisers on creativity as long as $\frac{v}{\gamma + r} > \frac{25\beta_{T+1}^2}{64\sum_{j=1}^{T-1}\beta_j}$.

History: The Collusive Recognition System

As soon as the agents “switched sides” and started working for the advertisers, they started to compete with each other more while struggling to converge on a systematic way to price their services. Although the creative-pitch contests described above occurred regularly to select the most creative agency for the job, the contests were not actually used to set the price competitively, as our model assumes. Instead, as we describe next, the agencies successfully colluded to demand 15% of media billings as the prize of winning the creative-pitch contests.

For approximately a century after the emergence of national advertising in the second half of the 19th century, the 15% rule was considered set in stone as the core element of an interrelated set of trade practices, collectively referred to as the *recognition system*. The publishers’ association “recognized” prospective agencies that met its standards. To be recognized, agencies were screened in areas of financial resources, advertising experience, types of accounts, moral standing, and bill-paying reputation. Agencies that passed the screening had their names disseminated to the publisher association’s members. While aiding in further stabilizing and legitimating the advertising business, these practices clearly inhibited market entry by new agencies. Such entry would clearly drive down the compensation of creativity-contest winners if the compensation were determined competitively.

Although we now think of the 15% commission as paid by the advertiser to the agency for the services rendered, the advertising industry of the late 19th century conceptualized the commission as a “discount” from the publishers’ list price that agencies were able to obtain when buying media space. For example, when the Pabst Brewery hired an agency to spend \$100,000 advertising the brand in national newspapers, the agency would produce media invoices totaling \$100,000 that included a \$15,000 “discount” pocketed by the agency and \$85,000 actually kept by the newspapers. Of course, such an arrangement invites side-dealing between the agency and the publishers. When the publishers included incentives on the side, these deals were called “kickbacks.”

Contemporaneously, large successful agencies took on a financing role, which made the publishers increasingly expect payments from agencies, not the advertisers themselves.¹¹ Thus, despite working for advertisers by developing and placing advertisements, the agencies maintained the illusion that they were getting paid by the publishers. This illusion also helped the agencies and publishers discourage price competition that would result if advertisers shopped around for advertising rates directly with the publishers. As early as 1893, the American Newspaper Association adopted a resolution stating that commissions in the form of discounts would be paid only to recognized advertising agents and not directly to advertisers (Trager 1992). Similarly, in 1902, the national magazines' trade association, the “Quoin Club,” advocated the use of the commission system.

In 1918, a year after its foundation, the American Association of Advertising Agencies (4A’s) was instrumental in further stabilizing the compensation practice at the standardized commission rate of 15% of the gross space rate: “It can be said that since 1918, on the testimony of agency men, agents have been a major factor in setting the rate of discount and have endeavored to make the rate they desired uniform for various classes of mediums. The figure set by agents as the desirable uniform rate, namely 15%, is one

¹¹ As early as the mid-19th century, advertising agents searched for profit sources beyond media commissions. Some purchased upfront space and profited from the differential between the rates they negotiated with publishers and the prices they charged advertisers. This practice turned financially strong agents into bankers that secured accounts by financing the advertising. Advertisers in turn were often more interested in the agent’s financing terms than the media price, and publishers looked for payment from agents and not from the advertisers.

which we said to reflect costs of doing business in 1918” (Haase 1934, p. 3). The fact that the publishers were no longer actually paying their agents as they used to in the past was not lost on industry observers: “The remuneration of the agent was not a payment from the publisher, or for services rendered to the publisher, but was a peculiar billing device under which the agent’s rate of payment was determined by agents according to their current (1918) costs of doing business” (Haase 1934, p. 27).

How profitable was an advertising agency that kept 15% of each dollar received from the advertiser? Sheldon (1925) shows that under the 15% commission rate, agencies actually retained only approximately 2% as profit, with the remaining 13% going to costs related to the servicing of their clients (see Table A1 in the Appendix for details). Others have shown agencies retained approximately 5% (Gundlach 1931; Goode 1933).

The *recognition system* limited price competition among agencies and enabled the bundling of the increasing range of services offered by independent full-service agencies. Agencies competed for business on the basis of service rather than price—increasing the importance of creativity and originality and eventually increasing the services provided. Leo Burnett emphasized the importance of "risk-taking creativity": "The advertiser wants ideas, needs ideas and is paying for ideas. We are going on the principle that every possible cent of income from an account should go into creative and productive efforts on that account." (Leo Burnett 1935, quoted in Tungate 2007, p. 68).

Agencies regarded the commission more as a royalty on sales than as an arbitrary service fee. Advocating this point of view, when Albert Lasker in 1925 took on the Lucky Strike account, he explained his position to his new client: “You’re spending \$600,000 or \$800,000 a year... and if in two or three years I can’t have this account to \$5 million, it can’t live. We’ll have to agree to do certain things, but when we do \$5 million, I never want to hear from you how much our commissions amount to. You won’t be paying them to us. We will have earned them ourselves” (Klaw 1956).

The above discussion shows both the newspaper and the agency associations were united behind the fixed 15% commission. Were they able to enforce the arrangement? In 1932, a survey of large advertisers was undertaken to investigate the “modern agency as advertisers see it” (Young 1933). All members of the

Association of National Advertisers (ANA), as well as the 100 leading magazine advertisers who were not ANA members, were surveyed. The response rate was 64%, resulting in a total of 223 advertisers. They represented 18% of total magazine advertisers and 59% of the total dollar volume of magazine advertising. Results indicated 97% of the advertisers paid the 15% commission rate and only 3% paid an annual or monthly fee against which all commissions were credited, and the fee exceeded total commissions.

As competition increased between advertising agencies, they searched for media types other than print to coordinate with the advertisers.¹² Radio advertising started in the 1920s and TV advertising in the 1940s,¹³ both initially in the form of shows sponsored by large advertisers; for example, the first “soap opera,” titled “Painted Dreams,” broadcast on the radio in the 1930s, was sponsored by Procter & Gamble’s line of laundry soap. Because these media (the networks and local radio and TV stations) were nurtured and financed by advertising, they also followed the *recognition system*, and their emergence failed to break up the collusive arrangement.

Advertisers challenged the recognition system as early as 1924, complaining that the refusal of some newspapers to give the 15% discount (i.e., the commission) directly to advertisers constituted a conspiracy in restraint of trade. Specifically, Vick Chemical Company claimed it was blacklisted from a list furnished to newspapers by advertising agencies, because of its policy of placing its own advertising. Other advertisers claimed to have established an agency under a separate name so that the discount would not be refused. For example, Carter Medical Company, as early as 1895, set up its internal advertising agency (under a separate name) because it foresaw the refusal of 15% to direct advertisers (*New York Times* 1927). On the flip side, in court, Firestone testified that the company would probably refuse to sell tires to a consumer at the wholesale rate (*New York Times* 1926), and the lawsuit was eventually dismissed in 1930.

To summarize, the *recognition system* was a collusive agreement between full-service advertising agencies and media owners to prevent direct dealings between advertisers and the media and to stop small

¹² Early media initiatives included outdoors advertising in the 1870s (posters, signs on rocks, buildings, and farms) and cable cars toward the end of the 19th century.

¹³ TV in 1949 was 1.3% of national advertising expenditures, but rapidly increased to 10.3% in 1953 and to 17.5% in 1958.

advertising shops and media independents from undercutting prices. Large advertisers (mostly national advertisers of branded products) and smaller non-member agencies continuously criticized the commission arrangement. The longstanding survival of the commission system indicates the power of the large advertisers and potential new entrants to the agency industry to change the compensation system was limited.

Model Solution: Incentives for collusion between agencies

When does our model predict that agencies are better off getting paid the old commission c than the expected creative commission δ , and thus have an incentive to collude against the advertisers? Consider period $T+2$, namely, the period after agents started working for advertisers instead of publishers. Suppose the inequality in Lemma 3 holds, so the publishers did not renew the contracts with the agents in $T+1$, and the agents are now working for the national advertisers. How high is their “net creativity commission” δ compared with the commission c they used to get? The creativity commission δ cannot be trivial (otherwise, the publishers would have hired them back; see Observation 1), but the range of parameters for which it is far below c is wide.

Proposition 1: As long as $1 > \frac{\delta}{c} > \frac{\Pi(\beta_{T+1})}{2\Pi(\beta_{T-1})}$, the agents earn a smaller net commission from working for

national advertisers and competing on creativity than the commission they used to get from publishers to acquire new local advertisers. In terms of model parameters, this sufficient condition reduces to

$$1 > \frac{16(1+kr\sigma^2)}{25(\gamma+r)} > \frac{\beta_{T+1}^2}{2\beta_{T-1}^2}.$$

Proposition 1 gives a sufficient¹⁴ condition under which the agencies are better off colluding than competing—conditions that were apparently operating in the real-world market throughout much of the 20th century. The proposition collects all the relevant parameters in a single inequality. The range of feasible

¹⁴ The proof also contains the necessary and sufficient condition, which is less elegant.

values is wide because it admits all values of $\frac{\delta}{c} \in \left[\frac{1}{2}, 1\right)$.

We now discuss the comparative statics of this key analytical result of our paper. Assuming the inequality in Lemma 3 is satisfied, the $1 > \frac{16(1+kr\sigma^2)}{25(\gamma+r)}$ inequality in Proposition 1 shows the agents are more likely to benefit from getting paid the old commission whenever costs of creativity γ are high—a very intuitive result. We guess that the cost of effective creativity γ rises over time, as the installed base of seasoned advertisers increases because being noticed in the ever-more-crowded field of advertisers is more difficult. Lemma 2 and Proposition 1 would thus imply a downward pressure on the creative commission and an increase in the profitability of collusion over time.

The benefit from collusion also increases when the costs of acquisition effort k and the associated variance σ^2 are low because these parameters increase the old commission. The only comparative static not immediately obvious from the inequality is the impact of risk aversion r : because $\frac{\partial}{\partial r} \frac{1+kr\sigma^2}{\gamma+r} \propto \gamma k \sigma^2 - 1$, risk aversion can both increase and decrease the benefit from collusion, and it is more likely to increase the benefits when either of the costs (acquisition or creativity) is high. Interestingly, the proportion of national advertisers v plays no direct role because v scales the total payoff regardless of which commission the agents receive.

Two observations are in order. First, note that in our model, advertisers are atomic, and the $2v\alpha_T$ advertising revenue is therefore just an amount of money spent by infinitely many tiny firms. Thus, the fact that the agency seems to be getting a cut of the advertising revenue after switching sides is not suboptimal per se. What is really happening is that the agency is getting compensated for the *increase* in the effectiveness of a given revenue. Importantly, the advertising budget is set by the advertiser, and so the existence of a commission does not imply the advertiser is giving the agency control over the revenue amount and offering a share of his revenue as some sort of misguided incentive. In other words, this model shows agents may not be paid enough after the switch, but not that the whole commission-based contract

type is necessarily wrong. Once large (non-atomic) advertisers emerge, we would expect them to exercise their market power and drive the agency compensation down, either by putting downward pressure on the commission amount or by asking for a volume discount in the form of a sliding-scale commission structure. As we discuss later in the paper, such price pressure from large advertisers did eventually lead to the demise of the collusive arrangement.

Second, our model abstracts away from new-advertiser acquisition after the agents go national. One way to incorporate it while preserving our qualitative results would be to assume the publisher can hire a lower-skill (lower β) agent to continue generating local business. As long as such an agent does not have the option of going national (maybe she does not have the right connections in other cities, maybe she is not part of the recognized set of agents), she does not have to be paid much.

History: Industry-standard 15% commission on media billings survives at least 20 years past the 1956 anti-trust ruling

Recall that the 1924 lawsuit challenging the recognition system as anti-competitive was eventually dismissed. The second time advertisers sued the publishers was in 1955, and this time, the federal antitrust authorities forced six trade associations representing both publishers and agencies to adhere to consent decrees, which

- (1) prevented explicitly fixing the agency commission at 15%;
- (2) required the media to contract with any entity with acceptable credit, not only the “recognized” agencies;
- (3) required all media to adhere to their published rate card; and
- (4) prevented the media from charging discriminatory rates to advertisers not using an agency. However, they could still withhold the “discount,” namely, commission, from such “direct” advertisers).

Although the 1956 consent decrees overall amounted to a victory for the advertisers, the victory was only partial as seen from point (4). The intended effect of the 1956 consent decrees was to formally dismantle the *recognition system*, but some of the decrees were almost impossible to enforce. Because compensation was still considered proprietary in client-agency relationships, deviations from points (1) and (3) were

rumored but difficult to prove. Advertisers and agencies receiving special treatment from publishers used intricate camouflage and maneuvering for the advertiser to avoid being called a “chiseler” and the agency to escape notoriety as a “cut-rate competitor” (Holland 1956). As a result of the unenforceability of some of the decrees, 15% media commissions persisted unaltered as the most widely utilized mode of agency compensation for at least 20 more years: the 1976 ANA survey revealed that 83% of responding agencies continued to get paid a commission based on media billings, and four out of every five of them continued to receive the 15% amount dating back to the 19th century.

3. How the Collusive Recognition System Eventually Disintegrated: Data from ANA Compensation Surveys

The previous section provided a model of agencies “switching sides” to work for advertisers, described the compensation challenges caused by this switch, and documented that agencies managed to collude with publishers to keep the pre-switch compensation scheme (a commission on media billings) and amount (15%). The collusive arrangement survived the emergence of new advertising media throughout the 20th century, as well as repeated anti-trust rulings. In this section, we first document how the collusive arrangement was gradually weakened by competitive pressure to reduce the commission amount throughout the 1980s, and then show how the core institution of compensating agencies based on media billings abruptly disappeared in the late 1990s after a new type of player—the media house—entered the agency industry.

Data: ANA surveys between 1932 and 2016

We analyze the series of compensation surveys conducted on a tri-annual basis by ANA since 1976. For an earlier period, we rely on the 1932 survey discussed earlier. ANA comprises around 500 advertisers from a cross section of U.S. industries, consisting of consumer and industrial manufacturers, retailers, service providers (insurance, communications, hotels, restaurants, utilities), and financial institutions. The ANA members include most large U.S. companies, such as Canon, IBM, American Express, Siemens, Warner

Brothers Studios, Verizon, and Carlson Hotels. We note that because almost every firm advertises, the membership of the ANA is obviously skewed in favor of large advertisers.

Each of the 14 surveys conducted between 1976 and 2016 used a random sample of the membership base. The size of the sample surveyed and the response rate varied: in the 1970s, the number of respondents was around 250; in the 1980s, around 160; in the 1990s and early 2000s, around 120; and in the more recent surveys, around 100 members. We do not have any reason to believe that the declining sample size is causing any sample-selection bias related to our metric of interest: the prevalence of billings-based and fee-based compensation schemes in the industry. Company executives (with titles ranging from advertising manager to marketing vice president) primarily responsible for advertising or marketing completed the questionnaires. Every survey in our sample asked about the details of compensation, and some of the surveys also asked additional questions relevant at the time. For example, surveys administered after 1988 also ask about performance-based incentives.

History: Slow erosion of the fixed 15% commission

As we explained in the previous section, the 1956 consent decrees had very little immediate impact on the type and amount of agency compensation. However, they did initiate a new trend of open discussions on compensation in the industry, allowing us to glimpse how all sides felt about the standard compensation structure at the time. The 4A's representing the agencies took the position that what had been working well should not be tinkered with, and they continued to advocate the continual use of the 15% commission as the best compensation method. In 1956, the 4A's president was quoted:

“Questions being raised had been asked in the early Nineteen Thirties and answered satisfactory at the time. The media commission method, while not perfect, was found then to be the most practicable one for maintaining the true and long-run interests of all advertisers and all media. In return for these commissions agencies will continue to render many services to media and the cause of advertising. Agencies develop new business for media, developing new advertisers and helping them (the media) grow.” (*New York Times* 1956a).

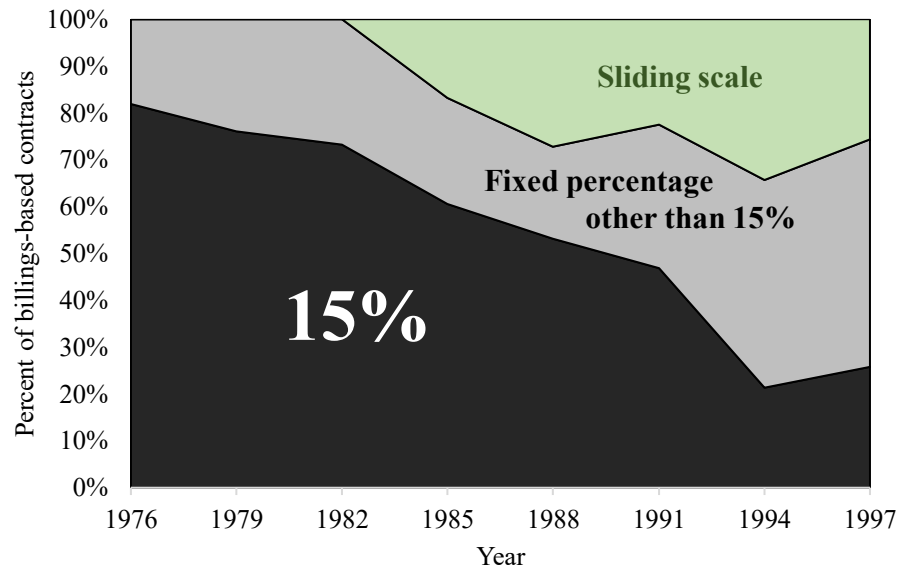
As part of the trend of open discussion, the 4A's publicly revealed its members' reported profit margins in 1959. The main message of the revelation was that costs were rising thanks to new value-added services,¹⁵ and non-compensated pitch development costs. If the agencies were to cut their commissions, they would need to cut these collateral services. Some agencies added a service charge of 17.65% to their procurement cost of these services, such as the cost of printing, artwork, graphics, or engravings, so that the agency would keep 15% of the overall invoice to the advertiser in line with keeping 15% of the media invoices. In 1959, agency charges distinct from media commissions amounted to about one third of their gross income. For large agencies, the amount was, on average, only 20%, whereas for small agencies (under \$10 million in billings), it exceeded income from commissions.¹⁶ This finding suggests smaller agencies charged and thus covered their costs for collateral services and possibly earned a profit on them, whereas larger agencies absorbed at least some of the costs of collateral services, because of the substantial media commissions they received. Meanwhile, advertisers continued to suspect the large agencies of receiving kickbacks from the media (i.e., unreported volume discounts), increasing their compensation further.

In summary, the 1956 consent decrees did not trigger an instant revolution in the industry's fixation on the 15% commission rate and the general idea of billings-based compensation. The 15% amount was so ingrained that it was even applied to other input charges incurred by the agencies. However, the idea of charging fees for some services took hold, and the 15% commission rate was broadly understood as compensating large agencies excessively (Appel 2017). In the late 1970s, the nature of agency compensation contracts finally started to noticeably shift: Figure 1 provides details about the types of billings-based compensation methods used over time: whereas 82% of advertisers using billings-based compensation paid their agencies the old fixed rate of 15% in 1976, only 21% did so by 1994.

¹⁵ "The developments of additional services was increasing rather than diminishing" (Gamble 1959, p.105). These additional services included research, marketing and merchandising counsel, public relations, consumer panels, Nielsen share of the market rating service, and even supplementing the client's field merchandising staff with agency employees. Research, for instance, was becoming more commonly used in 1959 (compared to the mid-1940s). Research was conducted to improve advertisements, determine the acceptability of products, test consumer preferences, find appealing themes, study the composition of markets, ascertain consumer needs and desires, and so on.

¹⁶ An account billing of \$300,000 to \$400,000 was considered borderline to no profit when only using commissions.

Figure 1: Breakdown of Billing-Based Programs 1976-97



The shift away from 15% was largely a shift to lower percentage amounts, as well as to sliding-scale rates that gave larger advertisers volume discounts, acknowledging the economies of scale and scope documented by Schmalensee et al. (1983) and Silk and Berndt (1993, 1994). In other words, the agencies experienced downward competitive pressure on the commission percentage throughout the 1980s and 1990s.¹⁷ One external shock that likely contributed to the downward price pressure was the well-documented contemporaneous shift in promotional budgets from advertising to trade promotions (Buzzell, Quelch and Salmon 1990). We propose that this weakening of the fixed 15% rate set the stage for the rather abrupt demise of the billings-based compensation, for reasons we describe next.

Theory: The impact of unbundling on agency compensation

Starting in the 1980s, alternatives to the full-service ad agency that both designed and placed the ads appeared on the scene. Specialized agents—“creative boutiques” and “media houses” that focus on only

¹⁷ One way to document the long-term downward price pressure is to compare the collusive 15% of media billings to the effective commission collected by today’s agencies. The 2016 provides the magnitude of the average agencies’ compensation as a percentage of the total advertising budget as 9.9% - a lot less than 15%. However, the lower percentage rate does not immediately imply lower overall compensation since media prices rose annually at a rate substantially exceeding that of the general inflation during the 1980s and 1990s (Arzaghi et al. 2012), and some leading agencies were perceived as over-compensated during the time of economic uncertainty and cost-cutting (Appel 2017).

one element of the advertising campaign—were becoming acceptable alternatives to full-service agencies. The appearance of media houses allowed advertisers of all sizes the option to unbundle their two advertising tasks (design and placement) to specialized agents, where the overall advertising strategy, the research, and the creative tasks continued to be performed by the full-service advertising agency (possibly with help from a creative boutique), while media planning and buying was delegated to the specialized media house. Horsky (2006) studies the unbundling phenomenon in detail and reports that by 2000, 58% of advertising dollars of advertisers with a budget greater than \$1 million were handled by media houses.

When an advertiser decides to place ads using a media house separate from the agency designing the advertising copy, the institution of paying either of the agencies a percentage of media billings becomes untenable for at least three reasons: First, the media billings become a private matter between the media house and the advertiser, so the creative agency does not routinely handle or even see the amount of money changing hands. The creative agency thus cannot verify that it is being paid a certain percentage of media billings. Second, the fiction of media “list prices” (long before undermined by “kickbacks” from publishers to agencies) gives way to the reality of the advertiser being quoted a media price that already contains a private profit margin to the media house. What matters to the advertiser is the cost of the media to him, not how the price breaks down into the media house’s margin and actual charges by the media. Finally, the unbundled arrangement clearly shows the creative agency is performing a service at a cost unrelated to the size of the advertiser’s advertising budget, making it difficult to argue that its compensation should be a percentage of that budget. Instead, compensating the creative agency the same way it compensates its other professional service providers makes sense—by a fee related to the labor and other costs incurred, possibly with a performance-based bonus to align incentives.

For all of the reasons described in the previous paragraph, the unbundling of agency services should result in a shift of agency compensation from billings-based to fee-based. Note the idea of fee-based compensation was not new in the advertising agency industry in the 1980s, but its implementation in a market dominated by publishers who were used to paying commissions was cumbersome and did not take

hold despite early attempts dating back to the 1960s.¹⁸ Fee-based compensation did not take over until the advertising budgets started flowing through media houses that cut private deals with publishers while creative agencies clearly performed a service unrelated to the size of the advertising budget. We now turn to validating these predictions by documenting that agency compensation indeed shifted to fee-based precisely after the unbundling became widespread.

History: Relationship between service unbundling and agency compensation

Figure 2 presents the co-evolution of compensation methods and unbundling between 1932 and 2016 based on the ANA surveys available to us. The black line with markers shows the percentage of billings-based contracts the advertisers reported. It clearly demonstrates both a long-term downward trend in the use of billings-based compensation (from about 80% in the 1970s to 3% in 2010) and a sharp drop over an only 10-year period from 61% in 1994 to 10% in 2003.¹⁹ The red line without markers shows the proportion of respondents reporting unbundling their creative and media services to separate agencies. These data show a sharp jump from 18% to 47% in just a single three-year interval between the 1991 and 1994 surveys.

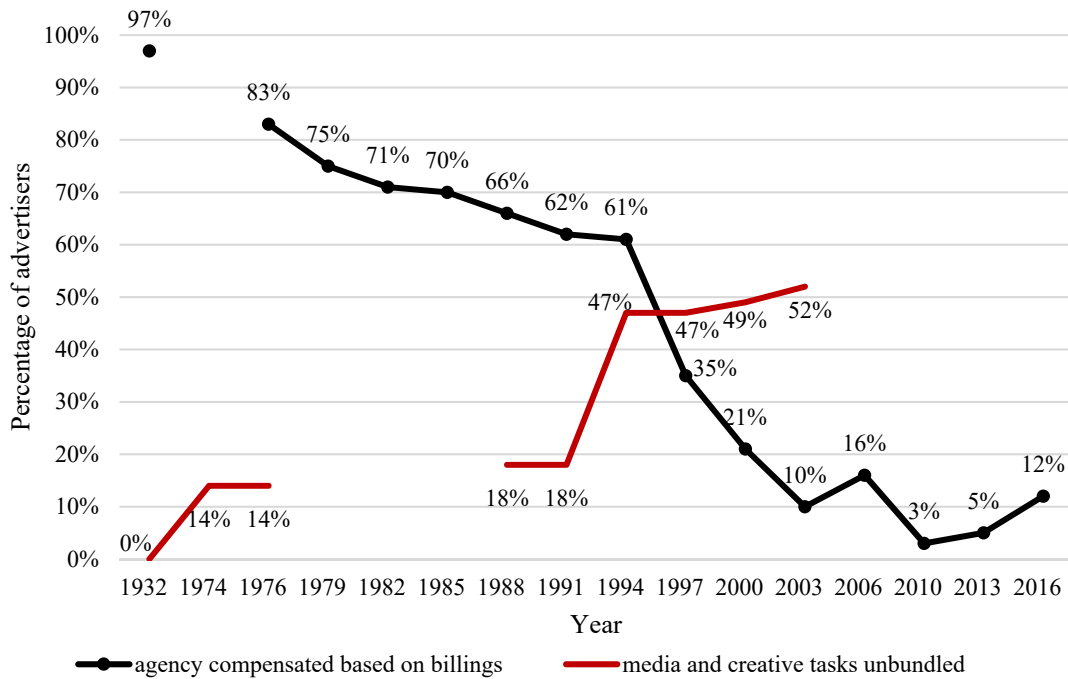
The sharp drop in the prevalence of billings-based compensation immediately follows the sharp rise in the proportion of advertisers unbundling their advertising services. Because we are not aware of other abrupt shifts occurring in the advertising industry in the mid 1990s, we propose that the unbundling caused the shift away from billings-based compensation in the sense of Granger. The gradual erosion of the 15% standard commission outlined in Figure 1 clearly set the stage for the unbundling's effect on the type of compensation: the focal 15% amount was relegated to a small minority of contracts by the early 1990s, so

¹⁸ A famous example of an early fee-based compensation contract is the 1960 arrangement between Shell Oil and its agency Ogilvy, Benson & Mather. The arrangement was a fee paid for one year's work determined as follows: the agency estimated its costs of direct and indirect salaries and overhead expense for the year and added 25% of those costs for profit. This scheme was actually difficult to implement, because the media was still counted on to pay the usual commissions to the agency, and the agency credited all commissions against the flat fee that was paid by the client: "The matter is delicate because neither client nor agency wants the media to get the idea that the 15% is a kickback and not a commission. A commission is part of the established business procedure. A kickback is something that could be withdrawn any time." This last point about "established business procedure" illustrates the difficulty of unilaterally shifting entrenched compensation schemes even by large players such as Shell Oil.

¹⁹ The small uptick in the use of billings-based compensation evident in the 2013 and 2016 surveys is entirely driven by the emergence of programmatic advertising, which is compensated on a commission basis.

total agency compensation became de facto negotiable, and expressing it as a percentage of media billings became a matter of otherwise meaningless convention. Once the unbundling shift occurred, just talking about total compensation and calling it a “fee” made sense. We propose that the unbundling was thus merely the straw that broke the commission camel’s back after more than a decade of advertisers’ pressure to reduce the commission rate. Even the agencies that continued being compensated based on the percentage of media billings experienced the 1991–94 unbundling shift in their compensations: the use of the 15% fixed rate dropped from 47% to 21% in the same time period (see Figure 1).

Figure 2: Co-evolution of Billings-based Compensation and Task Unbundling



Note to Figure: Agencies not compensated based on billings are compensated with a fee mostly related to their costs. The missing datapoints indicate years in which the question was not asked.

Note the compensation schemes took at least 10 years to fully shift from billings-based to fee-based. One reason for such a slow change is the fact the agencies sign multi-year contracts with advertisers, so not all advertisers were renewing all of their agency relationships in 1995. Another barrier to change was the deeply entrenched notion of a media commission as a standard industry procedure (see, e.g., the Ogilvy example in footnote 18).

4. General Discussion

This paper explains why advertising agencies used to be compensated with 15% of media billings for over 100 years even though they were working on creative ideas for the advertisers, and the incentives implied by a fixed media commission do not seem to be well aligned with the advertiser's goals. We model analytically and document empirically that the peculiar compensation scheme has roots in the early days of 19th-century newspaper advertising when the agencies worked for local publishers as commissioned salesmen of newspaper advertising space—a contract form standard in sales and easily rationalized by classic principal-agent theory. The agencies eventually switched sides and started working for the advertisers, specifically, the advertisers who had potential to go national (i.e., advertise beyond their local market). Our model explains this switch as an inevitable consequence of the emergence of national brand producers that wanted to advertise their good nationally and the agencies acquiring enough such potential national advertisers that bringing them to distant media markets became more profitable to the agencies than the compensation their local publisher was willing to offer. Note that at that point in time, national media did not yet exist and newspapers were localized. The dynamic that drove the agencies to switch sides operates in other principal-agent settings in which the agent accumulates client contacts during the course of her work for the principal and eventually breaks the contract to work for the clients instead. For example, consultants employed by management consultancies accumulate client contacts and often end up switching sides to work for one of the clients. The special feature of the advertising industry history we model is that the switch was synchronized and coincided with the emergence of national brands and advertising.

The fact that the agencies switched sides and started working for the advertisers at the end of the 19th century is only one piece of the puzzle that explains the seemingly suboptimal compensation scheme based on media billings. The second piece of the puzzle we document is the recognition system—a collusive arrangement between the agencies and the publishers that fixed the commission amount at 15% and erected barriers to entry into the agency industry. Our model characterizes the conditions under which the agencies first voluntarily switch sides and then want their old commission compensation back instead of receiving

competitive compensation based on the profit lift their creativity generates. We show these behaviors can happen under a wide range of parameters because cashing in the contact list at the time of switching sides reduces the outside options the agencies have, and so their old contracts are no longer available to them. Moreover, the national advertising market is much larger than the local market, so the agencies are willing to switch sides even when the expected commission available for creative work is much smaller than the commission they used to receive from publishers. In this case, colluding to get the old commission back is obviously beneficial even though the switch to work for the advertisers was voluntary.

The longevity of the collusive arrangement is surprising given the much more short-lived cartels that started at the same time, such as the Joint Executive Committee (JEC), analyzed by Ellison (1994). We can only speculate about the sources of this stability: unlike the JEC, which coordinated only among intermediaries (railroads shipping grain from suppliers in Chicago to customers on the East Coast), the advertising agency recognition system we document involves coordination of the intermediaries (the agencies) with the suppliers (the publishers). Moreover, advertising development and placement is a much more complex service than shipping grain, so the quality assurance embedded in the recognition system likely acted as an entry barrier. Perhaps the sources of the recognition system's longevity are in the features it shares with another long-lived collusive arrangement which we describe next.

The recognition system we describe has a parallel in the U.S. real estate industry, where the National Association of Realtors (NAR) has managed to sustain a collusive system that also involves a fixed commission and barriers to entry (Levitt, Syverson, and Ferreira 2008). Analogously for advertising agencies receiving a seemingly misguided commission on media billings, buyer agents in the U.S. receive a fixed commission on the price their client pays for the house. Also analogously, the agents started by representing sellers and only added buyer-side representation after the collusive Multiple Listing Service (MLS) was introduced. The NAR describes the MLS: "In the late 1800s, real estate brokers regularly gathered at the offices of their local associations to share information about properties they were trying to sell. They agreed to compensate other brokers who helped sell those properties, and the first MLS was born, based on a fundamental principal that's unique to organized real estate: Help me sell my inventory and I'll

help you sell yours.”²⁰ Besides not being able to spell “principle,” the NAR also signals the fundamental difficulty of compensating and motivating buyer’s agents—agents in charge of buying (spending money) rather than selling (getting revenue)—a difficulty both the real estate and the advertiser agents share. Conceptually, the classic principal-agent results should apply on that side of the market as well, giving the buyer’s agents an incentive to save money. However, such contracts are not common, at least to the extent of our knowledge. Instead, buyer’s agents receive fixed commissions on the money spent or flat fees—both poorly aligned with the goals of the buyer. We hope our paper stimulates future research into the effective contracts for buyer’s agents in both of these important markets and beyond.

As we point out, the eventual demise of the commission-based compensation method and its replacement by a cost-based fee resulted from the breakup of the “recognition system” cartel with the emergence of independent media houses, specializing in media buying, and the subsequent unbundling phenomenon. Paying the agencies for their creative work as a percentage of a media budget that they no longer purchased on behalf of the advertiser no longer made sense.

Our findings also have implications for today’s advertising market, which is undergoing a rapid shift to programmatic digital display advertising, with spending on display ads surpassing the spending on search ads since 2016. The 2013–16 uptick in billings-based compensation visible in Figure 2 is driven almost entirely by the fact that the intermediaries involved in programmatic advertising (e.g., media agencies, platforms, ad exchanges, and other aggregators) tend to be effectively compensated with a commission on the money passing through them. Often called “tech fees,” these commissions are ultimately paid by the advertiser, and whereas some are charged as commissions on the money spent, others are charged on a CPM basis (Beals and Elliott 2019). The ANA recently conducted at least three surveys regarding “media agency compensation” and discovered a lack of transparency in the way media agencies and other intermediaries spend the advertising budgets (Beals and Elliott 2019). So, history is repeating itself, and advertisers are again compensating their agencies on a commission basis instead of paying for added value.

²⁰ <https://www.nar.realtor/nar-doj-settlement/multiple-listing-service-mls-what-is-it>, accessed 7/17/2021. The NAR is currently being investigated by the DOJ for anti-competitive practices.

Much like the newspaper advertising market of the latter half of the 19th century, the programmatic display advertising market is currently in flux, evidenced by a lack of standardization and rapid changes in both the major players and market rules (Beals and Elliott 2019, Choi et al. 2020). As the market matures in the coming years, our findings suggest some major intermediaries will seek to collude with the publishers and retain the current lucrative billings-based compensation. Going forward, advertisers and regulators need to be careful about blindly embracing any sort of association of “certified” or “recognized” media agencies in the programmatic space, and they should be skeptical about any “standard contract” such an association promotes. Will the government have to step in to break up such an arrangement? Will regulatory effort be as unsuccessful as its analogue in 1956? Will some of the intermediaries eventually switch sides to work more explicitly for the advertisers, and start exclusively providing value-added creative services in return for fee-based compensation? Will such an unbundling in the digital advertising space lead to the demise of billings-based compensation? If any of the above happens, the history of ad agency compensation will be repeating itself.

References

- Appel, Marsha (2017). Defining Moments in Agency History | Evolution of Media Unbundling. Defining Moments in Agency History on aaaa.org
- Arzaghi M, Berndt ER, Davis J, Silk AJ (2012) The Unbundling of Advertising Agency Services: An Economic Analysis. *Review of Marketing Science* 10(1): 1-51.
- Association of National Advertisers, Inc. (1992,1995,1998,2001,2004,2007,2010,2013,2016) Trends in Agency Compensation. Association of National Advertisers, New York.
- Association of National Advertisers, Inc. (1976,1979,1983,1986,1989) Current Advertiser Practices in Compensating Their Advertising Agencies. Association of National Advertisers, New York.
- Basu AK, Lal R, Srinivasan V, Staelin R (1985) Salesforce Compensation Plans: An agency Theoretic Perspective. *Marketing Science* 4(Fall):267-290.
- Beals, David and Brittany Elliott (2019). Media Agency Compensation Practices. ANA.
- Bronnenberg Bart J., Dhar Sanjay K., Dubé Jean-Pierre (2007). Consumer Packaged Goods in the United States: National Brands, Local Branding. *Journal of Marketing Research* 44(1): 4-13.

- Buzzell, Robert. John Quelch and Walter Salmon (1990). The costly bargain of trade promotion. *Harvard Business Review* 68 (2). 141-9.
- Choi , Hana and Carl F. Mela and Santiago R. Balseiro and Adam Leary (2020), “Online Display Advertising Markets: A Literature Review and Future Directions”, *Information Systems Research* 31 (2), 556-575.
- Ellison, Glenn (1994). Theories of Cartel Stability and the Joint Executive Committee. *The RAND Journal of Economics* 25(1): 37-57.
- Fowler (1900) Fowler’s Publicity: An Encyclopedia of Advertising and Printing and all that Pertains to the Public-seeking side of Business (Boston).
- Fudenberg, Drew and Luis Rayo. (2019) Training and Effort Dynamics in Apprenticeship. *American Economic Review* 109 (111), 3780-3812.
- Gamble FR (1959) Advertising Agency Costs & Profits. *Harvard Business Review* 37 (November-December):103-112.
- Goode K (1933) Manual of Modern Advertising (Greenburg).
- Gross I (1972) The Creative Aspects of Advertising. *MIT Sloan Management Review* 14(1):83–109.
- Grossman S, Hart O (1983) An Analysis of the Principal-Agent Problem. *Econometrica* 51(January):7-45.
- Gundlach ET (1931) Facts and Fetishes in Advertising (Consolidated Book Publishers).
- Haase AE, Lockley LC, Digges IW (1934) Advertising Agency Compensation: Theory, Law, and Practice. Association of National Advertisers, New York.
- Hao Z (2005) Incentive-based compensation to advertising agencies: A principal–agent approach. *International Journal of Research in Marketing* 22(3):255-275
- Holland DR (1981) The Great Tradeoff. *Ad Age*. 51:41-42 (July 6).
- Holmstrom B (1979) Moral Hazard and Observability. *Bell Journal of Economics* 10 (Spring):74-91.
- Holmstrom B, Milgrom P (1987) Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica* 55(March):303-328.
- Holmstrom B, Milgrom P (1991) Multitask Principal-Agent Analysis: Incentive Contracts, Asset Ownership and Job Design. *Journal of Law, Economics and Organization* 7:24-52.
- Horsky D, Horsky S, Zeithammer R (2016) The Modern Advertising Agency Selection Contest: A Case for Stipends to New Participants. *Journal of Marketing Research* 53(5):773-789.
- Horsky, S., S. C. Michael, and A. J. Silk (2012), The Internalization of Advertising Services: An Inter-Industry Analysis. *Review of Marketing Science* 10: 1-37.
- Horsky S (2006) The Changing Architecture of Advertising Agencies. *Marketing Science* 25(4):367–383.

Jensen MC, Meckling WH (1976) Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics* 3(4):305-360.

Klaw S (1956) Advertising: The Battle of 15 Per Cent. *Fortune*. October:142-220.

Levitt, Steven D., Chad Syverson, and Fernando Ferreira. "Antitrust Implications of Outcomes When Home Sellers Use Flat-Fee Real Estate Agents." *Brookings-Wharton Papers on Urban Affairs*, 2008, 47-93.

New York Times (1926) Testifies on Advertising. (May 6), 13.

New York Times (1927) Charges Blacklisting on Its Advertising. (December 6), 52.

New York Times (1956a) Ad Men Support commission Pay. William M. Freeman (April 29), 61.

New York Times (1956b) News of the Advertising and Marketing Fields. (May 2), 40.

New York Times (1956c) Ad Group Studies Rate System, Decides against Hasty Changes. William M. Freeman (March 15), 45,56.

New York Times (1958) Advertising: Commission Hatchet is Buried. Carl Spielvogel (November 12), 60.

New York Times (1960a) Advertising: Shell and Ogilvy Try a Fee Instead of 15%. Robert Alden (November 16), 58.

New York Times (1960b) Advertising: Flat Fee Comes as a Surprise. Robert Alden (November 21).

Pope D (1983) *The Making of Modern Advertising* (Basic Books, New York).

Sheldon GH (1925) *Advertising Elements and Principles* (Harcourt, Brace and Company, New York).

Schmalensee R, Silk AJ, Bojanek R (1983) The Impact of Scale and Media Mix on Advertising Agency Costs. *The Journal of Business* 56(4):453-475.

Silk AJ, Berndt ER (1993) Scale and Scope Effects on Advertising Costs. *Marketing Science* 12(1):53-72.

Silk AJ, Berndt ER (1994) Costs, Institutional Mobility Barriers, and Market Structure: Advertising Agencies as Multiproduct Firms. *Journal of Economics & Management* 3(3):437-480.

Spake D, D'Souza G., Crutchfield T, and Morgan R (1999) Advertising Agency Compensation: An Agency Theory Explanation. *Journal of Advertising* 28(3):53-72.

Tungate M (2007) *Adland: A Global History of Advertising* (Kogan Page Limited).

Villas-Boas JM (1994) Sleeping with the Enemy: Should Competitors Share the Same Advertising Agency?" *Marketing Science* 13(1):190-202.

Wernerfelt, Birger, Alvin Silk and Shuyi Yu (2021) Internalization of Advertising Services: Testing a Theory of the Firm. *Marketing Science*. Published online in *Articles in Advance* 23 Aug 2021

Young JW (1933) Advertising Agency Compensation. In *Relation to the Total Cost of Advertising* (The University of Chicago Press, Chicago, IL).

Appendix

Standard preliminaries for tractability

Throughout, we assume linear contracts of the form $s+cX$, where X is the (either new or all, does not matter) advertising revenue of the publisher, s is salary, and c is commission.

Revenue linear in effort: The key simplification arising from the CARA utility assumption is that when the payoff follows a $\text{Normal}(\mu, \sigma^2)$ distribution, the certainty equivalent of a CARA agent with risk aversion

r is $\mu - \frac{1}{2}r\sigma^2$. So, if the advertising revenue is linear in effort, an additive error works: the output of agent's effort e is $X = \beta(e + \theta)$, where θ is a normal shock with variance σ^2 . Then, the agent's certainty equivalent

of the entire advertising revenue is $\underbrace{\beta e}_{\text{expected revenue}} - \underbrace{\beta^2 \frac{r\sigma^2}{2}}_{\text{risk premium}}$.

Quadratic cost of effort: Because the output is linear in effort, a convex cost is needed for concavity of the agent's objective function, and the quadratic specification is tractable in closed form, resulting in a linear first-order condition:

$$e^* = \arg \max_e \underbrace{s}_{\text{salary}} + \underbrace{c\beta e}_{\text{expected bonus}} - \underbrace{\frac{k}{2}e^2}_{\text{cost of effort}} - \underbrace{\frac{r\sigma^2}{2}(c\beta)^2}_{\text{risk premium}} = \arg \max_e c\beta e - \frac{k}{2}e^2$$

$$\text{FOC} : 0 = c\beta - ke^* \Leftrightarrow e^* = \frac{c\beta}{k} \quad (\text{IC})$$

At this optimal level of effort, the agent receives $s + \frac{(1 - kr\sigma^2)}{2k}(c\beta)^2$.

Observe: sign of $1 - kr\sigma^2$ is not a priori clear. Let's assume throughout that it is negative; that is, costs and/or risk aversion and/or variance are high enough. It is then clear that the agent does *not* benefit from increased marginal effectiveness of effort $c\beta$.

Risk-neutral principal: A publisher exists who maximizes the expected revenue minus expected compensation to the agent. Observe: the parameter r capturing agent risk aversion and parameter σ capturing the variability of the shock to effort appear only in the agent's preferences, only once and only multiplied together. Hence, let $V = r\sigma^2$ capture risk in general.

Proof of Lemma 1

The publisher solves

$$e^{**} = \arg \max_e \alpha + (1-c)\beta e - s \text{ subject to } c\beta = ke \quad (IC)$$

$$\text{and subject to: } s + c\beta e - \frac{k}{2}e^2 - \frac{r\sigma^2}{2}(c\beta)^2 = \underline{w} \quad (IR, \text{ which binds})'$$

where \underline{w} is the best change in wealth available to the agent outside of this contract. Note the IR is expressed in terms of the certainty equivalent flowing from the CARA assumption.

Substituting the base salary into the problem shows the publisher wants to maximize the total certainty equivalent subject to the IC:

$$e^{**} = \arg \max_e \alpha + \beta e - \frac{k}{2}e^2 - \frac{r\sigma^2}{2}(c\beta)^2 - \underline{w} \text{ subject to } c\beta = ke \quad (IC).$$

Plugging in the IC constraint and omitting the parts of revenue that do not depend on e results in

$$e^{**} = \arg \max_e \beta e - \frac{k}{2}e^2 - \frac{r\sigma^2}{2}(ke)^2$$

$$FOC: \beta - ke - rk^2\sigma^2 e = 0 \Leftrightarrow ke = \frac{\beta}{1+rk\sigma^2} \Leftrightarrow c = \frac{1}{1+rk\sigma^2}.$$

The publisher thus makes $\alpha - \underline{w} + \frac{\beta^2}{2k(1+kV)}$. QED Lemma 1

Proof of Lemma 2: Suppose the competitor has set her maximum creativity level to N . The focal agent's expected payoff from the contest, that is, the expected creativity commission p he receives when setting his level to some $M \leq N$, is

$$E(p) = E_\pi \left[\Pr(Y < \pi) E(\pi - Y | Y < \pi) \right] = E_\pi \left[\frac{1}{N} \int_0^\pi (\pi - y) dy \right] = E_\pi \left[\frac{\pi^2}{2N} \right] = \frac{M^2}{6N}.$$

The best response of the focal agency to the competitor is thus

$$M^{**}(\gamma, N) = \arg \max_M E(p) - (r + \gamma) \frac{Var(p)}{2} = \arg \max_M \frac{M^2}{6N} - (r + \gamma) \frac{M^3(3N - M)}{36N^2} =$$

$$= \frac{N}{8(r + \gamma)} \left[9(r + \gamma) + \sqrt{3} \sqrt{\frac{(r + \gamma)(27N(r + \gamma) - 128)}{N}} \right].$$

In a symmetric equilibrium, $M^{**}(\gamma, M) = M$, which solves to $M^{**}(\gamma) = \frac{24}{5(\gamma + r)} < M^*(\gamma)$. It can be

shown that when the competitor plays M^{**} , deviating above this value (a possibility not explicitly covered in the above derivation) is not profitable. The equilibrium level of creativity commission is

$E[p^{**}(\gamma)] = \frac{M^{**}(\gamma)}{6} = \frac{4}{5(\gamma+r)}$, and the net creativity commission retained by the agency after its costs

is $E(p^{**}(\gamma)) - (r+\gamma) \frac{Var(p^{**}(\gamma))}{2} = \frac{16}{25(\gamma+r)}$. QED Lemma 2

Proof of Lemma 3: $2\delta v\alpha_T > \Pi(\beta_{T+1}) \Leftrightarrow \frac{32v \sum_{j=1}^{T-1} \beta_j}{25(\gamma+r)k(1+kV)} > \frac{\beta_{T+1}^2}{2k(1+kV)} \Leftrightarrow \frac{v}{\gamma+r} > \frac{25\beta_{T+1}^2}{64 \sum_{j=1}^{T-1} \beta_j}$. QED L.3

Proof of Proposition 1: $1 > \frac{\delta}{c} > \frac{\Pi(\beta_{T+1})}{2\Pi(\beta_{T-1})}$ is just a combination of $\delta < c$ needed to prefer the old

commission to the expected creative commission in $T+2$, and $\frac{\delta}{c} > \frac{\Pi(\beta_{T+1})}{2\Pi(\beta_{T-1})}$, which is sufficient for

advertisers switching sides in $T+1$ because it implies $2\delta v\alpha_T > \Pi(\beta_{T+1})$. To see the sufficiency, note that

by definition of T , $c v\alpha_T > \Pi(\beta_{T-1})$, so $2\delta v\alpha_T > \frac{2\delta \Pi(\beta_{T-1})}{c}$, which in turn exceeds $\Pi(\beta_{T+1})$ as required

for advertisers switching sides when $\frac{\delta}{c} > \frac{\Pi(\beta_{T+1})}{2\Pi(\beta_{T-1})}$. Plugging in the definition of Π completes the proof.

The necessary and sufficient condition can be derived using Lemma 3, which shows

$2\delta v\alpha_T > \Pi(\beta_{T+1}) \Leftrightarrow \frac{v}{\gamma+r} > \frac{25\beta_{T+1}^2}{64 \sum_{j=1}^{T-1} \beta_j}$, and so $\frac{\delta}{c} = \frac{16(1+kV)}{25(\gamma+r)} > \frac{(1+kr\sigma^2)\beta_{T+1}^2}{4v \sum_{j=1}^{T-1} \beta_j}$. QED Proposition 1.

Table A1: Division of the 15% Commission (Sheldon 1925)

<u>Function</u>	<u>% of Advertiser's Expenditures</u>
Creative and contact	8.4
Research	1.3
Media, rate & checking	0.8
Accounting	0.7
Mechanical production	1.0
Administrative	0.5
Solicitation	0.3
Net Profit	2.0

Table A2: Model Notation

variable	variable description	variable type
c	agent's commission on media billings	contract settings in first epoch
s	agent's salary	
e	agent's effort of finding new advertisers	agent's decision in the first epoch
δ	agent's media commission for creativity	agent's payoff in second epoch
β	marginal effect of effort on revenue	free parameters describing agent's preferences and technologies
k	scale of the cost of effort	
r	agent's risk aversion (CARA)	
σ^2	variance of the random shock to effort	
ν	proportion of agents with national potential	
M	maximum lift due to creativity	
γ	scale of the agency's cost of increasing M	
V	$= r\sigma^2$, a composite risk parameter	additional useful notation
T	the time period in which agents switch sides	
t	time	background variables, including state variables
\underline{w}	outside option of agent	
α	installed base of seasoned advertisers	
θ	random shock to effort	
π	revenue lift due to creativity	
Π	"channel" profit from newly acquired advertisers net of risk and the cost of effort	