

Payment Scheme Self-Selection in the Credence Goods Market: An Experimental Study

Hernan Bejarano – CIDE and Economic Science Institute, Chapman University
Ellen P. Green – School for the Science of Healthcare Delivery, Arizona State University
Stephen Rassenti - Economic Science Institute, Chapman University

Given heterogeneity in expert behavior across payment schemes in credence goods markets, it becomes important to understand the consequences of payment scheme selection. To study the effect on customer well being of expert self-selection, we recruited subjects to participate in a real-effort credence good laboratory market. Experts were either randomly assigned or faced with the choice of three payment schemes: *fee-for-service*, *salary*, and *capitation*. We found that experts who selected *fee-for-service* payment resulted in customers with significantly worse outcomes in comparison with experts who had been randomly assigned to *fee-for-service*. In contrast, experts who selected *salary* payment did not change customer outcomes relative to those who were randomly assigned.

Highlights:

- Selection of fee-for-service and salary dominated over capitation.
- Self-selected Fee-for-service provided significantly more edits than randomly assigned.
- Self-selected Salary did not change the number of edits relative to randomly assigned.

Keywords:

Credence Goods, Self-Selection, Fee-for-service, Capitation, Payment, Sorting

1. Introduction

The characteristic trait of a credence good market is that customers are never aware of whether the services they were provided were of the type or quality that they needed; they only observe their marginal utility of receiving those services. On the contrary, the service providers, i.e., experts, have better information regarding both assessment and appropriate action, presenting them with a nearly ideal environment to place their interests ahead of their customers' (Emons 1997). For example, physicians have better information regarding the necessary treatments for their patients both before (considering reported symptoms) and after (accounting for examination and test results) an appointment. Given this informational asymmetry, traditional solutions for aligning customer and expert's preferences, such as punishment and reward incentives, monitoring, liability, verifiability, and reputation, may not be available or effective (Brosig-Koch, Hennig-Schmidt et al. 2013, Dulleck and Kerschbamer 2006, Green 2014, Hennig-Schmidt, Selten et al. 2011). Customer well being would therefore be largely determined by experts' intrinsic motivations. Understanding how monetary incentives and expert preferences influence sorting is of importance to employers and policy makers. Our study places subjects in a credence goods environment in order to examine how expert self-sorting by payment scheme affects customer well being.

Our experiment builds on two independent branches of the credence-goods literature: payment schemes and self-selection. Research on payment schemes in credence goods markets is mixed. In particular, some empirical studies utilizing naturally occurring data found that expert behavior was independent of payment schemes (Crosson, Propper et al. 2001, Devlin and Sarma 2008), while others found a significant change in behavior from payment schemes (Gaynor and Gertler 1995). The latter studies found that experts altered their actions as self-interested profit maximizers (Matsaganis and Glennerster 1994, Robinson 2001, Robinson, Shortell et al. 2004). Systematic literature reviews of payment schemes generally conclude that there is not enough supporting evidence to recommend policy changes (Gosden, Forland et al. 2001, Sørensen and Grytten 2003).

Disparities in empirical studies are likely caused by the nature of the data. Payment schemes are difficult to study in context because naturally occurring data oftentimes does not allow the researcher to control for expert's unobservable characteristics (e.g., altruism and risk aversion), distinguish between implicit and explicit incentives (e.g., wages v. promotion and tenure), adapt for measurement error, and finally, implement exogenous policy changes. These hurdles motivate laboratory experiments. Experiments can control for selection in payment schemes and also condition experts and customers payment functions to induce different utility functions (Smith 1976). Furthermore, the experimental environment isolates the relationship between customers and experts, allowing researchers to focus on a single question. Following this line of research, there is a general consensus that suggests that expert behavior is influenced by payment scheme. In their experimental studies, Green (2014) and Hennig-Schmidt (2011) both showed that experts were motivated by payment scheme as well as how their actions impacted their customers. In their experiments, experts deviated from their profit-maximizing

actions to improve their customers' well being. However, the deviation from profit-maximizing behavior varied across payment schemes. Hence, even without self-selection productivity differentiation by payment scheme holds true (Corgnet, Hernán-González et al. 2014).

Given disparities in actions across payment schemes, it is important to understand expert self-selection. Experiments on sorting have found that self-selection by payment scheme interacts with experts' skills, and performance, thereby affecting productivity in labor markets (Cadsby, Song et al. 2007, Dohmen and Falk 2011, Macpherson, Prasad et al. 2014). More specifically, workers who were given the opportunity to select their payment scheme in a labor market were more productive with piece rate than flat rate (as expected), but another interpretation of this result is that piece rate payment can attract more productive workers (Lazear 1996). The key difference between past self-selection literature and our experiment is the introduction of the credence goods market. This market allows for intrinsic motivations and information asymmetries that may have a significant impact on expert actions.

To elicit the effects of expert self-selection in a credence goods market, we used a proofreading task that was imbedded in a two treatment, two by three experimental design. Treatment 1 varied the assignment mechanism (self-selection or random), while Treatment 2 varied the payment scheme (*salary*, *capitation*, or *fee-for-service*). Under each payment scheme, experts faced a moral dilemma or tradeoff between maximizing earnings and delivering customer-centered service. Under *salary*, experts faced a tradeoff between minimizing their effort (making fewer edits) and increasing the well being of their customers. In contrast, under *fee-for-service*, experts could maximize their payment by providing as many edits as possible, irrespective of their correctness or appropriateness. Finally, in *capitation* experts faced a tradeoff between maximizing the number of customers serviced by minimizing effort for each customer. In this way, we can observe both the expert's preference for moral dilemma circumstances, and the strength of that moral dilemma as manifested by their subsequent actions.

We found several relevant results. First, selection of *fee-for-service* and *salary* dominated over *capitation*. Secondly, on average we found no significant change in the customers' well being when comparing expert's behavior in randomly assigned with those in the self-selection treatment. However, disaggregating by payment scheme, we observed significant differences in expert actions under the *fee-for-service* payment scheme. More specifically, experts who selected *fee-for-service* provided significantly more edits (services) than those who were randomly assigned. This increase included a significant number of harmful edits made by these experts, while the number of beneficial edits remained constant. In comparison, experts who selected *salary* payment did not change the number of edits (beneficial or harmful) relative to those who were randomly assigned *salary* payment. Given these results, customers could expect to be better off when matched with salaried experts.

To the best of our knowledge, this is the first study to investigate effects of self-selection by payment scheme in a credence goods market with a real effort task (see Brosig-Koch et al. for a chosen effort task). Credence goods markets have a particularly important impact on the welfare of our society. The healthcare industry alone accounted for 17.4% of our GDP in 2014

(Hartman, Martin et al. 2015). Hence, even small market failures, such as the overprovision of services, would result in large losses in customer welfare.

2. The Experiment

2.1. Experimental Task

Our experimental task was first introduced in Green (2014) to study the effect of various physician payment schemes on physician performance. An advantage of the design is that it can be readily interpreted as a representation of the relationship between any expert and customer. Therefore, we have broadened the interpretation of Green's dual-principal agent game to all credence goods markets and adopted the credence terminology throughout our paper.

As in Green (2014), our experiment was implemented in two phases. In phase I, subjects were recruited to play the roles of customers. The customers proofread 10 unique essays in 50 minutes. Each essay had 10 spelling and/or typographical errors. At the start of the session, the customers were endowed with \$25. For each error the customers failed to identify, \$0.25 was taken away from their initial endowment. Customers were paid according to their performance, however they were not informed which errors they were able to correctly identify and which errors remained uncorrected. After all customers completed phase I, essays from each customer were selected for proofreading by the experts in phase II. The customer's lack of expertise in correcting the essays allowed us to endogenously generate a demand for the credence good. That is, each customer needed the expert's proofreading services to recover some of their endowment in phase II.

In phase II, subjects were recruited to act as experts and re-proofread the essays on behalf of their customers. The errors the customers were able to identify were removed from the text between phases I and II, which left only the errors that needed to be corrected in the essays presented to the experts. There were 125 errors that needed to be corrected. Each expert was presented with 4 rounds of essays, each round comprised of 10 essays. The essays presented to all experts in each round were the same, but the errors requiring editing changed from round to round. Experts had a maximum of 7 minutes to complete each round; however, if the expert was satisfied with her edits she was able to proceed to the next round before the 7 minutes were up. Between rounds each expert received a report of her previous round and cumulative results, including her number of correct edits, her number of incorrect edits and her change in aggregate earnings. In summary, experts were presented with 40 essays over 4 rounds with a total of 125 errors to detect in a maximum of 28 minutes.

2.2 Credence Goods Markets in the Laboratory

Each of the 125 errors the expert needed to correct was highlighted in turquoise, but for each highlighted error two additional sections of text were also highlighted that did not contain actual errors. Therefore, 375 sections of text were highlighted in turquoise across the 40 essays, 125 sections contained actual errors and 250 were without errors. The expert was informed of

this ratio, but was free to demarcate any highlighted section of text as containing an error, which allowed for her over-provision of services,

Figure 1:
 The **root** to the stop **was** familiar from countless hours of riding bikes, roller-blading, and the like, but today, it seemed completely **foreign** to me.

Green (2014)

Figure 1 is an excerpt from one of the essays presented to the experts in phase II. In this example, the customer failed to identify the spelling of the word “root” as incorrect in phase I (the correct spelling in this context is “route”). The other highlighted words, “was” and “foreign”, did not need to be amended, but were highlighted to allow the experts to provide unnecessary service to their customers. Note, the customer’s imbursement in phase II was dependent on quality, while the expert’s imbursement was dependent on quantity.

Table 1: Expert and Customer Payment

Phase I	
Customer	Initially endowed with \$25.00, decreased by \$0.25 for each error left unidentified
Phase II (Randomly Assigned and Self-Selection)	
Expert	
Fee-for-service	\$0.20 for each edit
Capitation	\$0.625 for each essay edited
Salary	\$25.00 at the end of the experiment
Customer	
Harmful Edits	\$0.15 each
Beneficial Edits	-\$0.05

Treatments in the experiment varied by assignment mechanism (self-selection or random), while Treatment 2 varied the payment scheme (*salary*, *capitation*, or *fee-for-service*). Table 1 provides a summary of the payments schemes. The *self-selection* treatment allowed the experts to select among the payment schemes used in the *fee-for-service*, *capitation*, and *salary* treatments. In the *fee-for-service* treatment, experts were paid \$0.20 per edit to the text that they made, regardless of correctness. In the *salary* treatment, payment was independent of edits made to the text; experts were paid \$25 to participate in the experiment. In the *capitation* treatment, experts were paid \$0.625 for each essay in which they edited at least one highlighted section of the text, regardless of correctness. *Fee-for-service* and *capitation* parameters are set such that experts who provided only necessary services obtain the same earnings as those under *salary*.¹

¹ The payment systems differed from Green’s dual-principal agent game. In Green (2014) payments were determined such that if subjects behaved in their customer’s best interest, they would earn the same amount under each treatment.

The customer's well being was impacted by the quality of the expert's edits in phase II. The impact of expert's actions on customer well being was constant across treatments. The customer's well-being decreased by \$0.05 for each section of the text that was incorrectly amended by the expert and increased by \$0.15 for each section of the text that was correctly amended. There was no change in the customer's well being for sections of the text that remained unchanged. For instance, in example 1 above, if the expert correctly edited "root" and incorrectly edited "was" the payments to the customer and expert were as follows:

- The customer would receive \$0.10 (\$0.15 from "root" and -\$0.05 from "was")
- If the expert was in the *salary* treatment, her salary would be unaffected by the edits made.
- If the expert was in the *fee-for-service* treatment, she would receive \$0.40 for the edits (\$0.20 for each edit).
- If the expert was in the *capitation* treatment, she would receive \$0.625 for having made at least one edit (correct or not) on behalf of this customer.

2.3 Implementation

The experiment was conducted at the laboratory of the Economic Science Institute (ESI) at Chapman University. A total of 167 undergraduate students participated. Experimental sessions were conducted between May 2014 and May 2015. Essays were borrowed from Green's original experiment. Essays were written by an English professor and contained errors that he commonly saw in his college students' work. Once the essays were created, they were assessed for difficulty in a principles of microeconomics class at Virginia Tech and calibrated accordingly.

Proofreading in phase I was conducted in ESI laboratory classrooms. Following their task, phase I students were given the option to be partially reimbursed for their missed errors based on the performance of a second group of students (phase II). A total of 8 students participated in phase I, from which an average of 5 essays from each student were chosen to create the panel of 40 essays to be used across all treatments in phase II. On average subjects earned \$13.53 from their performance in phase I.²

A total of 159 undergraduate students participated in phase II of the experiment.³ Phase II was conducted at computer stations within the ESI laboratory. Students were separated into cubicles. Printed instructions were provided for the students to read on their own for ten minutes. At the end of the 10 minutes, the experiment coordinator read the same instructions out loud. Next, to ensure task understanding, students completed a quiz on either their assigned payment scheme (in treatments *fee-for-service*, *capitation*, and *salary*) or a quiz on all three payment schemes (in the *self-selection* treatment). After 10 minutes, the answers to the quiz were

²Despite multiple attempts by the authors, a large majority of the subjects in phase I chose not to collect their reimbursements from phase II.

³One subject's data was discarded in our analysis because she restarted previously completed proofreading.

evaluated. This process continued in shorter intervals until all subjects correctly answered all quiz questions.

Once all subjects had answered the quizzes correctly, subjects began proofreading essays via the computer interface. Essays were presented in 4 rounds. Each round consisted of a panel of ten essays. Subjects had a maximum of 7 minutes to proofread and correct the essays in each round. The computer interface allowed subjects to navigate back and forth between the 10 essays within the round. Students did not have access to any other software in the duration of the experiment. However, if the students chose to do so, they could select to finish the proofreading task at any point in the experiment. On average, in the phase II sessions lasted 1.5 hours and phase II subjects earned \$29.44 in addition to the \$7 show up fee.

Table 2: Subject Count by Experimental and Treatment Status

Player	Payment	N	Totals
Customers		8	8
Experts			
Randomly Assigned	Capitation	25	105
	Fee-For-Service	39	
	Salary	41	
Self-Selection	Capitation	2	53
	Fee-For-Service	21	
	Salary	30	
Total			166

3. Results

Our results are consistent with previous literature that compares the *fee-for-service*, *salary*, and *capitation* payment schemes (Green, 2014). Experts in *fee-for-service* provided more services at a lower quality than the *salary* and *capitation* payment schemes under both the *self-selection* and the *randomly assigned* treatments. Table 2 presents the distribution of subjects under each treatment and payment structure. When selection was allowed, we observed that two payment schemes, *fee-for-service*, *salary* were preferred over *capitation* by the experts. We observed that out of 53 in the *self-selection* treatment, 30 (56%) chose *salary*, 21(42%) *fee-for-service*, and 2 (3%) *capitation*. The remainder of this section describes our 9 key results, delineated by overall *self-selection* effects (Results 1-5) and payment scheme specific *self-selection* effects (Results 6-9).

Result 1: Self-selection induced a greater overall number of edits.

Chart 1 demonstrates the first result of our experiment. Experts made 145.1 edits on average in the *self-selection* treatment, while only 118.4 in the *randomly assigned* treatments. This significant increase of 25% (Fligner-Policello Rank U=-1.454 p-value = 0.07303) implies

that on average under *self-selection* experts provide more services than when they are *randomly assigned* a payment scheme. However, in our experimental design, experts who provided more edits did not necessarily provide a higher overall quality of service. Hence, we need to disaggregate edits by harmful and beneficial to better understand the overall impact of self-selection. Here, beneficial edits are defined as those that are correct. Harmful edits can be either necessary but incorrectly edited or unnecessary.

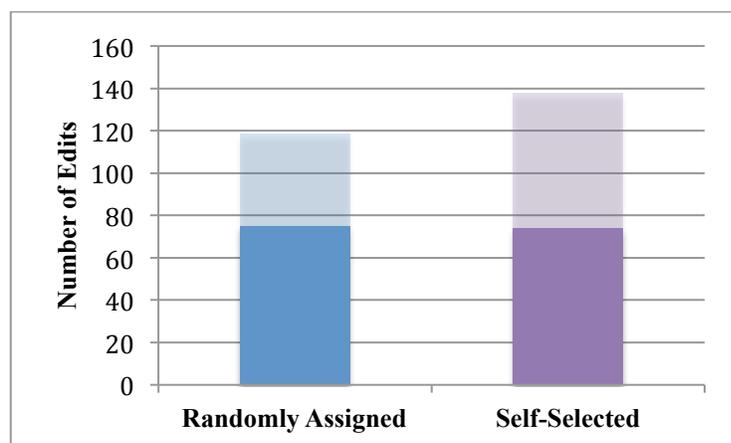


Chart 1 Beneficial Edits and Harmful Edits: The translucent portion of the bar represents harmful edits; the opaque section of the bar represents beneficial edits.

Results 2: Self-selection did not affect the overall number of beneficial edits.

We observe almost no difference in the average number of beneficial edits provided between the *self-selection* and *randomly assigned* treatments. Experts provided on average 72.2 beneficial edits in the *self-selection* treatments, while those in the *randomly assigned* treatments provided 75.0 beneficial edits. The observed increase was not statistically significant (Fligner-Policello Rank $U=0.047$ $p\text{-value} = 0.48113$). Hence, experts acting in self-selected payment schemes do not seem to increase or decrease the average number of beneficial edits provided.

Result 3: Self-selection induced a greater overall number of harmful edits.

Experts in the *self-selection* treatment provided an average of 73.0 harmful edits, while those who were *randomly assigned* provided only 43.4 harmful edits. This constituted a significant increase of 68% in the average number of harmful edits (Fligner-Policello Rank $U=-1.405$ $p\text{-value} = 0.07997$). In light of Result 1, we see experts who *self-select* increase the number of edits provided relative to experts *randomly assigned*, but Results 2 and 3 indicate that this difference is driven by an increase in harmful rather than beneficial edits.

Result 4: Self-selection did not affect the overall impact on the customer's well being.

In our experimental design, the customer's payoff or well being is affected by the quality of the expert's edits, both harmful and beneficial. Almost no difference was observed in average customer well being across the *self-selection* and *randomly assigned* treatments. The *randomly-assigned* experts increased their customers' well being by \$9.08 on average. This was larger but not statistically different than *self-selection* experts who increased their customer's well being by \$7.18 on average (Fligner-Policello Rank $U=0.655$ $p\text{-value}=0.25630$).

We evaluate expert effort in two ways: 1) The net impact on customer well being as calculated by payoff generated by the parameters of the experiment, or 2) The ratio of beneficial-to-total edits. The latter criterion simultaneously evaluates both quantity of edits provided and the characteristics of the edits (i.e., beneficial or harmful), and seems to provide a more representative measure of the expert's actions in the proofreading task.

Result 5: Self-selection did not affect the overall ratio of beneficial-to-total edits.

Aggregating across all payment schemes, we did not observe a significant difference between the ratios of beneficial-to-total edits which were 0.70 for *self-selection* experts and 0.78 for *randomly assigned* experts (Fligner-Policello Rank $U=-1.049$ $p\text{-value}=0.14700$).

Results 1-5 summarize the effects that the *self-selection* treatment had on the overall provision of the credence goods. The following 5 results isolate the effects that the *self-selection* treatment had on the performance within the *fee-for-service* and *salary* payment schemes.⁴ Table 3 reports Fligner-Policello rank comparisons between *randomly assigned* experts and *self-selection* experts by payment scheme.

Result 6: Self-selection did not affect the number of beneficial edits under salary or fee-for-service.

Consistent with Result 2, the average number of beneficial edits provided by experts who selected either *fee-for-service* or *salary* payment schemes was not significantly different than the number provided by experts randomly assigned to these payment schemes. This result is not surprising given that none of the payment schemes increased experts' earnings with *beneficial* edits. Table 3 row 2 presents the mean number of beneficial edits under each payment scheme disaggregated by the *self-selection* and the *randomly assigned* treatments. The average number of beneficial edits under the *fee-for-service* payment scheme in the *self-selection* treatment is smaller than in *randomly assigned* experts (68.4 and 77.8, respectively); the opposite is true under salary (74.9 and 77.5, respectively). However, neither of these differences is statistically significant.

⁴ Only two experts in the *self-selection* treatment chose the *capitation* payment scheme, therefore, we were unable to explore the effects sorting had on this payment structure.

Table 3: Experimental Results by Payment Scheme and Assignment Mechanism

	Fee-for-service			Salary		
	Randomly Assigned	Self-Selection	Fligner-Policello	Randomly Assigned	Self-Selection	Fligner-Policello
1. Total Edited						
Mean	175.8	238.2	0.02042	81.7	84.6	0.25562
SD	(106.1)	(119.8)		(29.4)	(30.6)	
Median	123.0	283.0		88.0	95.5	
2. Beneficial (B)						
Mean	77.8	68.4	0.19464	74.5	74.9	0.35581
SD	(28.8)	(36.3)		(27.9)	(32.1)	
Median	88.0	74.0		81.0	83.5	
3. Harmful						
Mean	98.1	169.8	0.00490	7.2	9.7	0.17335
SD	(123.2)	(144.3)		(4.9)	(11.2)	
Median	23.0	178.0		6.0	7.5	
4. Impact						
Mean	\$6.80	\$1.80	0.04225	\$10.80	\$10.70	0.37561
SD	(9.6)	(11.8)		(4.1)	(5.0)	
Median	10.7	6.9		11.8	12.2	
5. Beneficial/Total Edited						
Mean	0.56	0.40	0.00900	0.90	0.90	0.44783
SD	(0.3)	(0.3)		(0.1)	(0.2)	
Median	0.80	0.40		0.90	0.90	
N	39	21		41	30	

Data reported describes the sum of variable over the four rounds.

Fligner-Policello tests the hypothesis that the *randomly assigned* and *self-selected* medians are the same within payment schemes. P-values are reported.

Result 7: Self-selection did not affect the number of harmful edits under Salary, but self-selection increased the number of harmful edits under fee-for-service.

In Result 3, we noted that the increase in edits in the *self-selection* treatment was largely driven by the increase in number of harmful edits provided. Table 3 row 3 reports the average number of harmful edits provided by payment scheme. Here we see that experts who selected *fee-for-service* provided on average 73% more harmful edits than experts randomly assigned to this payment scheme. In contrast, experts who selected *salary* payment did not significantly change the number of harmful edits produced compared with experts who were randomly

assigned. Hence, the increase in harmful edits observed in the *self-selection* treatment is driven by harmful edits made by those who selected *fee-for-service* payment.

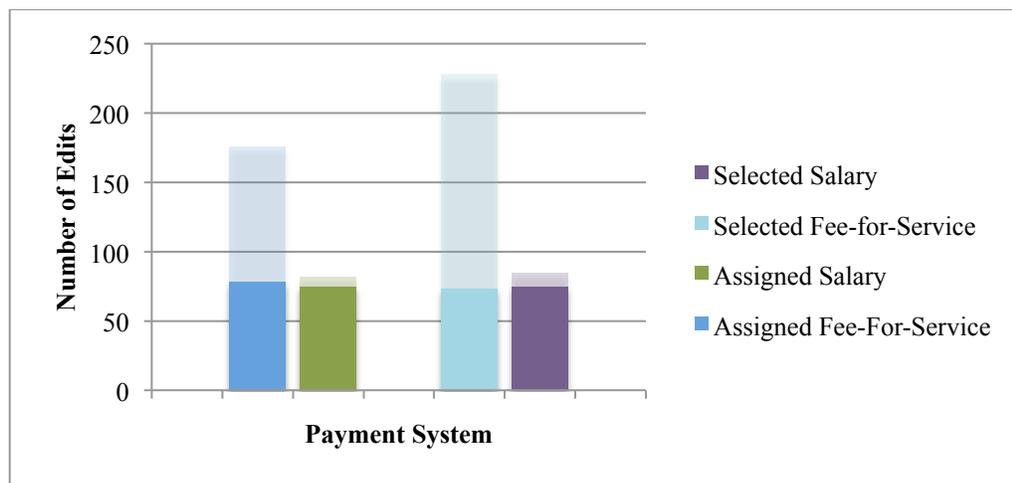


Chart 2 Beneficial Edits and Harmful Edits by Payment: The translucent portion of the bar represents harmful edits; the opaque section of the bar represents beneficial edits.

Result 8: Self-selection did not affect the expert’s impact on the customer’s well being in the salary payment scheme, but self-selection did negatively impact the customer’s well being under the fee-for-service treatment.

Table 3 row 5 reports that an expert who selected *salary* payment would provide approximately \$5 more benefit to a customer than an expert who selected *fee-for-service* payment. This can also be seen in Chart 2, which uses a bar chart to show the number of harmful and beneficial edits made by experts who were assigned or selected salary payment. Result 4 indicated that *self-selection* did not significantly increase or decrease the customers’ expected well being. This means that a customer who randomly chose an expert could expect a similar outcome whether the expert was assigned a payment scheme or selected it herself. In spite of this, Results 7 and 8 reveal that knowing whether the payment scheme of her expert was assigned or self-selected provides valuable information to the customer. In general, a customer’s expected well being is higher with experts paid via *salary* than *fee-for-service*; however, under *self-selection* this distinction is even greater.

Result 9: Self-selection did not affect the ratio of beneficial-to-total edits under salary payment, but self-selection did affect the ratio of beneficial-to-total edits under fee-for-service payment.

Our analysis of the ratio of beneficial-to-total edits yields two interesting findings shown in Row 5 of Table 3 and Chart 3. Chart 3 reports the number of edits that were beneficial and harmful disaggregated by payment scheme. The average ratio of beneficial-to-total edits is greater for *salary* than for *fee-for-service* payment. The average ratio of beneficial-to-total edits for experts paid by *salary* is 0.90, in both the *self-selection* and the *randomly assigned* treatment.

However, the average ratio of beneficial-to-total edits under *fee-for-service* payment is 0.56 for *randomly assigned* experts which is significantly different than 0.40 for *self-selection* experts (Fligner-Policello Rank U=0.009 p-value =.0090). Row 5 Table 3 shows that while the introduction of *self-selection* did not affect the average ratio of beneficial-to-total edits for experts on *salary*, it does seem to drive experts paid by *fee-for-service* to provide a lower ratio of beneficial-to-total edits.

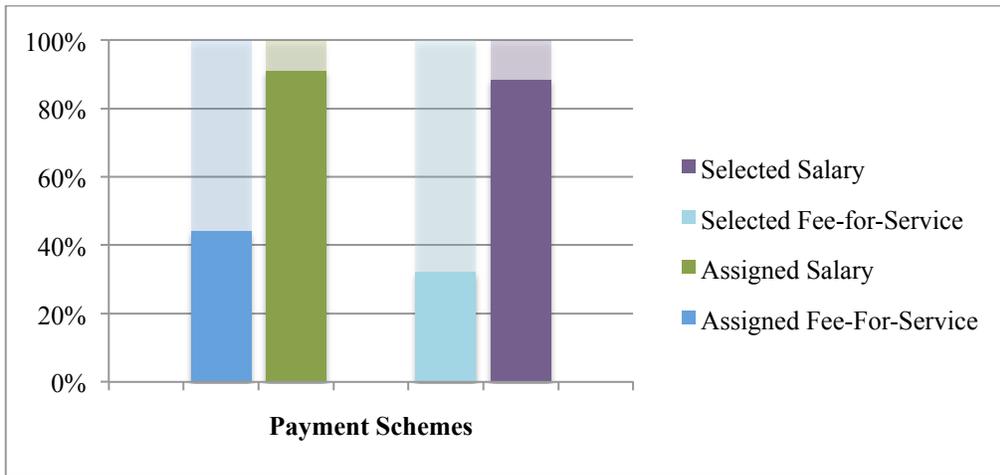


Chart 3 Percentage Beneficial and Harmful Edits by Payment: The translucent portion of the bar represents harmful edits; the opaque section of the bar represents beneficial edits.

5. Conclusions

In our experiment experts were either randomly assigned or allowed to select their payment scheme in a credence goods market. Selection is a fundamental characteristic of the labor market and has been investigated in traditional industries (Cadsby, Song et al. 2007, Dohmen and Falk 2011, Macpherson, Prasad et al. 2014), but ours is the first real effort experimental study to investigate selection in the credence good environment. Our experiment reveals the importance of incorporating self-selection into credence goods models, as it impacted expert's actions and thereby customer outcomes under the two most selected payment schemes, *fee-for-service* and *salary*. More specifically, experts who selected *fee-for-service* payment provided more harmful edits than those who selected into *salary* or those who were randomly assigned to *fee-for-service*. This resulted in lower customer well being and a lower ratio of beneficial-to-total edits. Hence, self-selection exacerbated the asymmetric information problems inherent to the credence goods market in the *fee-for-service* payment scheme. Conversely, selection had no significant impact on expert behavior under the salary payment scheme.

These findings and experimental design provide a solid foundation for future studies to test external validity and robustness across different credence goods environments. For instance, in our experiment, experts were constrained to choose between different payment schemes, but were not able to enter different markets (i.e., different professions). In contrast, in naturally occurring credence markets experts have outside options and hence, selection out of a profession may occur due to an interaction between payment scheme and profession. That is, the subset of

future experts in a profession may be dependent on profession preference, payment schemes present, and ability to select among them. Future experiments should provide outside options for experimental experts to address the effects of self-selection on the formation of expert subgroups and the resulting behavior. If the finding is robust to different experimental parameters and environments, it could mean that a credence good customer would benefit from knowing both her expert's payment scheme and how her expert's position was obtained (i.e., through selection and matching mechanisms). Furthermore, knowing that self-selection may directly influence an expert's profession of choice is an important finding for employers and policy makers designing rules that influence the number and characteristics of doctors in hospitals, teachers in schools, and other experts to their profession.

7. References

- Brosig-Koch, J., et al. (2013). "How Effective are Pay-for-Performance Incentives for Physicians?-A Laboratory Experiment." A Laboratory Experiment (April 2, 2013).
- Cadsby, C. B., et al. (2007). "Sorting and incentive effects of pay for performance: An experimental investigation." Academy of Management Journal **50**(2): 387-405.
- Corngnet, B., et al. (2014). "Why real leisure really matters: Incentive effects on real effort in the laboratory." Experimental Economics **18**(2): 284-301.
- Crosson, B., et al. (2001). "Do doctors respond to financial incentives? UK family doctors and the GP fundholder scheme." Journal of Public Economics **79**(2): 375-398.
- Devlin, R. A. and S. Sarma (2008). "Do physician remuneration schemes matter? The case of Canadian family physicians." Journal of health economics **27**(5): 1168-1181.
- Dohmen, T. and A. Falk (2011). "Performance pay and multidimensional sorting: Productivity, preferences, and gender." The American Economic Review: 556-590.
- Dulleck, U. and R. Kerschbamer (2006). "On doctors, mechanics, and computer specialists: The economics of credence goods." Journal of Economic literature: 5-42.
- Emons, W. (1997). "Credence goods and fraudulent experts." The RAND Journal of Economics: 107-119.
- Gaynor, M. and P. Gertler (1995). "Moral hazard and risk spreading in partnerships." The RAND Journal of Economics: 591-613.
- Gosden, T., et al. (2001). "Impact of payment method on behaviour of primary care physicians: a systematic review." Journal of Health Services Research & Policy **6**(1): 44-55.
- Green, E. P. (2014). "Payment systems in the healthcare industry: An experimental study of physician incentives." Journal of Economic Behavior & Organization **106**: 367-378.

Hartman, M., et al. (2015). "National health spending in 2013: growth slows, remains in step with the overall economy." Health Affairs **34**(1): 150-160.

Hennig-Schmidt, H., et al. (2011). "How payment systems affect physicians' provision behaviour—an experimental investigation." Journal of health economics **30**(4): 637-646.

Lazear, E. P. (1996). Performance pay and productivity, National bureau of economic research.

Macpherson, D. A., et al. (2014). "Deferred compensation vs. efficiency wages: An experimental test of effort provision and self-selection." Journal of Economic Behavior & Organization **102**: 90-107.

Matsaganis, M. and H. Glennerster (1994). "The threat of 'cream skimming' in the post-reform NHS." Journal of health economics **13**(1): 31-60.

Robinson, J. C. (2001). "Theory and practice in the design of physician payment incentives." Milbank Quarterly **79**(2): 149-177.

Robinson, J. C., et al. (2004). "The alignment and blending of payment incentives within physician organizations." Health Services Research **39**(5): 1589-1606.

Sørensen, R. J. and J. Grytten (2003). "Service production and contract choice in primary physician services." Health Policy **66**(1): 73-93.