

Do control questions influence behavior in experiments?

Catherine Roux · Christian Thöni

Received: 12 June 2012 / Revised: 14 February 2014 / Accepted: 5 March 2014 /
Published online: 22 March 2014
© Economic Science Association 2014

Abstract Outcomes and strategies shown in control questions prior to experimental play may provide subjects with anchors or induce experimenter demand effects. In a Cournot oligopoly experiment we explore whether control questions influence subjects' choices in initial periods and over the course of a repeated game. We vary the framing of the control question to explore the cause of potential influences. We find no evidence for an influence of the control question on choices, neither in the first period nor later in the game.

Keywords Control questions · Experimenter demand effects · Anchoring · Experimental design

JEL Classification B41 · C72 · C91

1 Introduction

Control questions which test the subjects' understanding of instructions prior to experimental play are standard procedure in experimental economics. In these questions, one cannot avoid showing specific combinations of strategies and

Electronic supplementary material The online version of this article (doi:[10.1007/s10683-014-9395-y](https://doi.org/10.1007/s10683-014-9395-y)) contains supplementary material, which is available to authorized users.

C. Roux
FGN-HSG, University of St.Gallen, Varnbuelstrasse 19, 9000 St.Gallen, Switzerland
e-mail: catherine.roux@unisg.ch

C. Thöni (✉)
University of Lausanne, UNIL-Dorigny, 1015 Lausanne, Switzerland
e-mail: christian.thoeni@unil.ch
URL: <http://sites.google.com/site/christianthoeni>

outcomes which may influence subjects' play, especially in the initial phase of the experiment.¹

Although of considerable methodological interest, the influence of control questions on subjects' behavior has not yet been examined systematically. We investigate the effects of quantities shown in the control question on immediate actions and over the course of 20 periods of a repeated Cournot game. We see mainly two channels by which control questions may affect behavior. First, they can induce experimenter demand effects. Experimenter demand effects refer to changes of subjects' behavior due to cues about what constitutes appropriate behavior (Zizzo 2010). They have been shown to influence behavior in various experimental environments (Zizzo and Flemming 2011; Bardsley 2008). Second, outcomes presented in control questions may serve as an anchor. Anchoring in decision making occurs when individuals rely too heavily on an initial piece of information, judging by its actual importance to the situation at hand. Psychological research on how anchoring influences individuals' beliefs, value judgement and information processing is abundant and suggests that anchoring affects a variety of numerical estimates (see Furnham and Boo (2011) for a literature review).

We introduce a treatment variation to shed light on these two explanations. In half of our sample, we inform the subjects of the fact that the situation in the control question is randomly generated. Subjects then know that the situation is not deliberately chosen by the experimenter and therefore contains no information about appropriate behavior. Thus, when the information about the random process is given, we should not observe any influence of the control question on behavior if the influence is driven by experimenter demand effects.² If, however, anchoring is the mechanism through which control questions influence choices the treatment variation should not matter.

2 Experimental design and procedures

Similar to the design of Huck et al. (2004) $n \in \{2, 4, 6, 8\}$, subjects play a repeated Cournot game with linear demand and constant marginal costs of 2. We use a loaded frame explaining the game in terms of firms and production quantities. There are 20 identical periods with partner matching. The firms simultaneously and independently choose their quantity q_i from $[0, 74]$ at an increment of .1. The lower and upper limits of the choice set are not actively communicated. Only when the subjects try to enter a quantity outside the interval, they are informed about the

¹ “[Q]uizzes [i.e., control questions] run the risk of giving unintended cues about your intentions to your subjects” (Friedman and Sunder 1994, p. 52).

² Zizzo (2010) distinguishes between social (wanting to please the experimenter) and cognitive (using all information received from the experimenter as cues about what constitutes optimal behavior) experimenter demand effects. In principle, our treatment variation shuts down both channels because a random number conveys no information about either optimal behavior or the experimenter's preferred outcome. If, however, subjects, rightly so, guessed that the purpose of the experiment is to find out how they react to random numbers, a positive effect could well be explained by social experimenter demand effects. We deem this very unlikely because the instructions were focused on the Cournot game while the control question was only a minor preparatory step.

range of admissible quantities. Price is determined by the market demand $p = \max\{74 - Q, 0\}$ where $Q = \sum_i q_i$ is the total quantity produced in the market. At the end of each period, players learn all individual quantities and profits in their market. Firm i 's profit in the stage game is $\pi_i = pq_i - 2q_i$. The Nash equilibrium is $q^n = 72/(n + 1)$, i.e., 24, 14.4, 10.3, and 8 for our markets with 2, 4, 6, and 8 firms.

Prior to the first period, subjects are confronted with a control question in which they have to calculate their payoff for a combination of the own quantity (q_i^{cq}) and the average quantity of the other players (\bar{q}_{-i}^{cq}). The wording is:

Please consider the following situation:

Your production quantity	10
Average production quantity of the 3 other firms	8.2
What is your profit in this situation (rounded to integers)?	_____

Both quantities displayed are generated randomly from the uniform distribution in $[0, 74/n]$.³ We adjust the interval for group size to avoid situations with very high quantities that result in negative payoffs.⁴ To facilitate the calculation, subjects have access to a profit calculator which allows them to determine the payoff consequences of hypothetical combinations of their quantity and the average quantity of the other firms. The control question is the only occasion when the subjects observe a specific combination of quantities provided by the experimenters.

Our empirical strategy is to test whether the situation presented in the control question has predictive power for the quantity decisions, formally, we look for the slope of the function $q_i = f(q_i^{cq})$. To explore potential explanations for an influence of the control question on behavior we introduce a treatment variation: in *Standard* we present the subjects with the control question as shown above. In *Random* we add the following information between title and situation: “(The situation in this control question was generated purely randomly)”.⁵ The allocation to treatments was randomized within sessions such that half of the subjects were in *Standard* and half were in *Random*.

What might be the size of the effect? While we are not aware of any studies providing estimates for the strength of experimenter demand effects, there is a number of studies reporting quantitative results on anchoring effects. The estimates for f' vary considerably. Ariely et al. (2003) finding values between .2 and .5, while most subsequent replication studies find either smaller effects or no effect at all. Maniadis et al. (2014) review this literature and find an average effect of about .13.⁶

³ q_i^{cq} is rounded to an integer, and \bar{q}_{-i}^{cq} is rounded to a multiple of .1.

⁴ Negative payoffs are still possible if both draws are close to the upper bound. In our experiment, this situation did not occur.

⁵ See the online appendix for screenshots of the control-question stage and the instructions.

⁶ We use the results reported in Tables 1 and 2 in Maniadis et al. (2014) to calculate the imposed slopes. Many studies do only report two classes of anchors (low and high) but not the particular value of an anchor. In these cases we calculate the slope for the intermediate value of the range of anchors. Maniadis

We have observations from 214 subjects. The numbers of observations are 30 (17), 64 (35), 72 (32), and 48 (23) for the markets with 2, 4, 6, and 8 firms respectively; the numbers of observations in *Random* are shown in parantheses. The sessions were run in the lab at the University of St.Gallen during February and October 2012 and were programmed in z-Tree (Fischbacher 2007).⁷ Recruitment was done with ORSEE (Greiner 2004). Subjects were randomly allocated to computer terminals such that they could not infer with whom they would interact. During the entire experiment communication was not allowed. We provided written instructions which informed the subjects of all the market features, and we read out loud the most important points. Profits were calculated in ‘Guilders’, and 1000 Guilders were worth 1.10 CHF for the duopolies. We adapted the exchange rate for the four, six and eight-firm markets such that Nash earnings were equalized across markets. Payments consisted of a show-up fee of 10 CHF plus the sum of the profits over the course of the experiment. Sessions lasted for about 105 minutes, and the average earning was about 37 CHF (~ 39 USD). The subjects were undergraduate students in economics, business, international studies and law at the University of St. Gallen, and had not participated in oligopoly experiments before.

3 Results

We start by examining the effect of control-question quantities on output choices in the first period. We pool the data from the two treatments. Figure 1 plots the relation between q_i^{cq} and q_i in the first period separately for the different markets. For orientation, we add a dashed 45° line, the location of observations where subjects choose exactly their control-question quantity. The solid line is an OLS regression line, and we indicate the 95 % confidence interval for the slope (shaded area). The plots show no systematic relationship. In markets with four and eight firms we estimate a positive slope, while in markets with two and six firms we estimate a negative slope. All slope coefficients are insignificant at $p > .16$.

Result 1 The quantities shown in the control question do not seem to have an effect on subjects’ quantity choices in the first period of the experiment.

Insignificant results must, of course, be interpreted with great caution. If the empirical setup does not offer sufficient power to detect the effect the lack of significant results cannot be interpreted in a meaningful way. To address this concern, we use a Monte-Carlo simulation and answer the following question: how likely is it to observe our parameter estimates if the true effect of the control-question quantity on output decisions was of a specific value? In particular, we use

Footnote 6 continued

et al. (2014) report an estimate of .14, while other replications studies range from $-.5$ to $.46$. We calculate a weighted average for all replication studies with the number of observations as weights.

⁷ After the subjects had played the repeated Cournot game, the experiment continued with a Cournot game with punishment. The results are discussed in Roux and Thöni (2013).

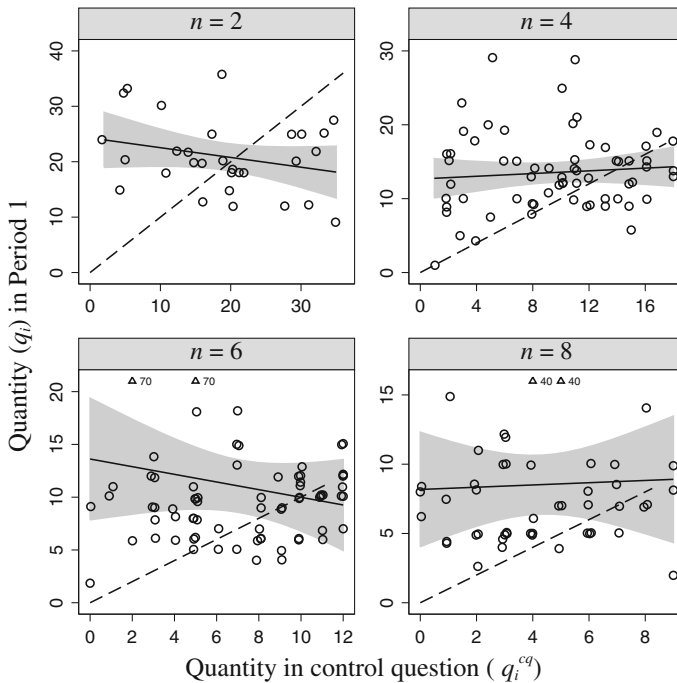


Fig. 1 Quantity chosen in the first period of the Cournot game in relation to the quantity in the control question, separately for the number of firms in the market. *Dashed line* is the 45° line, *solid line* shows predicted quantities from an OLS regression, *shaded area* shows the 95 % confidence interval for the slope of the regression line. *Dots* are jittered. Outliers with very high quantities are indicated as *triangles* and not drawn to scale; the *small numbers* denote the q_i of the outliers

the values $\alpha = .2$, $.1$, or $.05$ (recall that the estimates for the size of anchoring effects range from 0 to $.5$ with an average of $.13$, see footnote 6). We randomly draw quantities q_i^r from a normal distribution with mean and standard deviation as observed in our experiment. For each n we draw the same number of observations as in our experiment. We use our 214 control-question quantities and calculate hypothetical quantities $q_i^h = q_i^r + \alpha(q_i^{cq} - q_i^r)$.⁸ We ran four times 2,000 regressions with q_i^h as the dependent and q_i^{cq} as the independent variable. We compare the distribution of coefficients for q_i^{cq} with the slope coefficients of the four regression lines in Fig. 1. For example, for $n = 2$ we estimate a slope of $-.176$ in our sample. In the 2,000 regressions on simulated data with $\alpha = .2$ only a single estimate resulted in a coefficient lower than $-.176$, suggesting a p value of $.0005$. For $n = 4$ we observe a slope of $.092$. This is obviously much more likely to happen if the true slope is $\alpha = .2$, resulting in a p value of $.163$. The estimated p values for six and eight firm markets are $p = .24$ and $p = .379$. Using Fisher's method to combine the p values from the four independent tests results in $p = .0004$, i.e., our estimates are very unlikely to be observed if the true effect of the control question on quantities is

⁸ We also restrict the hypothetical quantities to feasible quantities, $q_i^h \in [0, 74]$.

.2.⁹ Lower values of α are more likely to result in the estimates we observe. In case of $\alpha = .1$, we get an overall p value of .016, while for an effect size as small as $\alpha = .05$, our setup results in a weakly significant p value (.089). Alternatively, we may identify the effect in the entire sample using dummies to control for the number of firms. In this model, we measure a coefficient of $-.115$ for q_i^{cq} . Assuming an effect size of .2, we find zero out of 2,000 simulated estimates to result in a lower coefficient, with an effect size of .1 we find six estimates with a lower coefficient, suggesting a p value of .003, an effect size of .05 results in $p = .035$. Taken together, our results strongly suggest that in Cournot games the effect of the control-question quantity on behavior—should it exist—is smaller than the typical estimates of anchoring effects found in the literature.

Still it is premature to conclude that experimenter demand or anchoring effects are unimportant. The analysis so far does not take into account that subjects are presented with two quantities in the control question, their own and the others' average quantity. It might be that the effect runs via the latter, e.g. if subjects tend to play best response to the others' quantities shown in the control question. We use OLS estimates explaining q_i by both q_i^{cq} and \bar{q}_{-i}^{cq} for each number of firms to address this issue. All eight coefficient estimates are insignificant with $p > .2$.

While we do not find evidence for an effect in the whole sample, we could still find a difference between the two treatments, *Standard* and *Random*. To investigate this possibility we ran OLS estimates where we interact both control-question quantities with a dummy for the treatment variation. Table 1 shows the results. With one exception, all interaction effects as well as the dummy for *Random* are far from significance, the exception being the interaction between the treatment dummy and the other firms' quantity for $n = 2$. Given that the suggested effect is absent in the other three regressions, we believe that this simply reflects spurious correlation and should not be interpreted as an effect. The insignificant F tests and the low adjusted R^2 also suggest that the explanatory power of the five variables is basically zero.

Result 2 We find no difference between the treatments *Standard* and *Random*. Informing the subjects about the fact that the situation in the control question is randomly generated does not seem to influence the effect of the control question on behavior in the game.

So far, we only tested for overall effects of the control question quantity on choices. Research on anchoring suggests that not all anchors are equally important for choices. In particular, Sudgen et al. (2013) show that anchoring effects are restricted to situations where the anchor has a plausible value in the context of the decision situation. In our game, this could mean that the control question is only affecting behavior if the situation shown is 'reasonable' in the sense that it leads to a good outcome for the subject. We perform a median split with respect to the profit implied by the combination of q_i^{cq} and \bar{q}_{-i}^{cq} shown in the control question. Similar to the estimates shown in Table 1, we interact the dummy for the median split with the

⁹ According to Fisher's method the p values from k independent tests can be aggregated to a joint test, where the test statistic $-2 \sum_{i=1}^k \ln(p_i)$ is χ^2 distributed with $2k$ degrees of freedom. See Fisher (1958, p.99f).

Table 1 OLS estimates for quantity in period 1

	$n = 2$	$n = 4$	$n = 6$	$n = 8$
q_i^{cq}	-0.146 (0.196)	-0.070 (0.203)	-0.465 (0.613)	0.144 (0.292)
$Random \times q_i^{cq}$	0.060 (0.236)	0.287 (0.273)	0.164 (0.871)	-0.310 (0.606)
\bar{q}_{-i}^{cq}	0.420 (0.247)	-0.079 (0.175)	-0.539 (0.586)	-0.027 (0.175)
$Random \times \bar{q}_{-i}^{cq}$	-0.625** (0.264)	0.165 (0.255)	0.844 (0.684)	-1.252 (1.096)
$Random$ (D)	7.107 (7.989)	-5.076 (4.152)	-7.082 (10.666)	9.796 (6.964)
Constant	17.486** (7.237)	15.240*** (2.922)	18.146* (9.373)	6.451*** (1.657)
F -test	2.0	0.4	0.4	0.5
Prob > F	0.114	0.861	0.869	0.761
R^2 adjusted	0.222	-0.041	-0.032	0.035
N	30	64	72	48

OLS estimates for individual quantity (q_i) in period 1 with robust standard errors. Independent variables are the quantities shown in the control question (q_i^{cq} and \bar{q}_{-i}^{cq}), a dummy for the treatment variation $Random$, and interactions

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

two control-question quantities. None of the interaction effects with q_i^{cq} are significant, and the point estimates have different signs across group sizes. For the interactions with \bar{q}_{-i}^{cq} , we observe weakly significant coefficients for $n = 4$ and 6. However, the two coefficients have the opposite sign. Alternatively, we perform a median split with respect to how close q_i^{cq} is to q^n , the Nash-equilibrium quantity. Here, all interaction effects are insignificant, and the point estimates do not suggest a systematic direction. Taken together, we do not find evidence for stronger effects of control-question quantities on behavior for reasonable anchors.¹⁰

So far, we only considered data from the first period of the game. Given that we observe no systematic effect in the first period, we do not expect to find one in later periods either. We run the same estimates as reported in Table 1 for all 20 periods with robust standard errors clustered on group level and controlling for period. With one exception all coefficients of control-question quantities and interactions are insignificant $p > .1$. The significant coefficient refers to q_i^{cq} in the estimate for $n = 6$ (coefficient: -0.105 ; $p = .020$). Alternatively, we model the dynamics of the quantity choice by adding lagged q_i and \bar{q}_{-i} as explanatory variables. All control-question quantities are insignificant ($p > .1$), but in the estimate for $n = 2$ we

¹⁰ The effect of the control question on behavior could depend on cognitive abilities (Bergman et al. 2010). While we do not have an independent measure of cognitive ability, we can use the time spent in the control question stage as a proxy for how quickly the subjects understood the task. Running regressions as in Table 1 with a variable for time instead of the treatment variation does not suggest that subjects spending more time in the control question are more reactive to the control question.

observe a weakly significant effect for q_i^{cq} (-0.046 , $p = .094$). As the coefficients are insignificant or negative, we find no support for a positive relation between the control-question quantity and the quantity choices of the subjects over the course of the 20 periods.

4 Discussion and conclusion

Control questions typically present subjects with a specific combination of strategies and outcomes. This may influence behavior, either because subjects perceive the situation shown as cue about desired behavior or because of unconscious anchoring effects. We investigate the influence of control questions on behavior in a repeated Cournot game. As a treatment variation, we inform the subjects about the fact that the situation in the control question was randomly generated. We find no evidence for an influence of the quantities presented in the control question on subjects' output choices, neither in the first period nor later in the game. Moreover, our treatment variation does not alter this finding. Our results point in the same direction as the recent experimental findings in the literature on anchoring effects which suggest that these effects are typically weaker than the ones reported in early studies (Maniadis et al. 2014).

A word of caution is in order here. While our results suggest the absence of experimenter demand or anchoring effects induced by control questions, they should not be misread as being evidence against the importance of experimenter demand effects per se. These effects may well be introduced by factors other than control questions such as the experimenter's authority, framing of the instructions or the lack of experimenter-subject anonymity (Barnettler et al. 2012; Zizzo 2010).

To what extent our results carry over to other experimental games or to designs which make use of a series of control questions remains an open question. We can think of the following factors that may play a role: (i) the complexity of the game environment, and (ii) the set-up of the control questions. First, the derivation of the Nash equilibrium in Cournot oligopoly games is fairly complicated, relative to simpler games like Dictator or Ultimatum games. Existing experimental evidence suggests that familiarity with the task, experience, knowledge and less uncertainty in decision making weakens the influence of anchors (Sudgen et al. 2013; Wilson et al. 1996; Chapman and Johnson 1994; Tversky and Kahneman 1974). Moreover, it is also likely that control question induced experimenter demand effects are stronger in more complex games. The reason is that, in simple environments such as a Dictator game, subjects who want to please the experimenter immediately know what to do if they assume the researcher is looking for pro-social behavior. In more complex games this is much less obvious, and we think that subjects are most likely to use the control-question situation as an indication for what the experimenter might want to see.

Second, in many experiments subjects have to answer a series of control questions which systematically explore the strategy space. While having a number of questions most likely reduces anchoring effects, it is not clear what happens to the strength of experimenter demand effects. On the one hand, cues pointing in opposite directions may weaken demand effects (Sitzia and Zizzo 2011) but, on the

other hand, with several questions, subjects have to think harder about all possible outcomes which may make demand effects stronger.

Our conjecture that complex games are most vulnerable to the influence of control questions as well as the investigation of multiple control questions is left for further research. To generate further evidence on these issues should not be difficult. Given that most of the experimental projects use control questions anyway, the data comes almost as a free lunch. We would like to encourage experimenters to implement randomly generated strategy combinations for the control questions and to look out for influences on experimental play.

Acknowledgments We thank the Editor and two anonymous referees for very helpful comments and suggestions. Financial support from the Richard Büchner-Stiftung is gratefully acknowledged. The usual disclaimer applies.

References

- Ariely, D., Loewenstein, G., & Prelec, D. (2003). Coherent arbitrariness: stable demand curves without stable preferences. *Quarterly Journal of Economics*, 118(1), 73–105.
- Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics*, 11(2), 122–133.
- Barmettler, F., Fehr, E., & Zehnder, C. (2012). Big experimenter is watching you! Anonymity and prosocial behavior in the laboratory. *Games and Economic Behavior*, 75(1), 17–34.
- Bergman, O., Ellingsen, T., Johannesson, M., & Svensson, C. (2010). Anchoring and cognitive ability. *Economics Letters*, 107(1), 66–68.
- Chapman, G. B., & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making*, 7(4), 223–242.
- Fischbacher, U. (2007). Z-tree—zurich toolbox for readymade economic experiments. *Experimental Economics*, 10(2), 171–178.
- Fisher, R. A. (1958). *Statistical Methods for Research Workers* (thirteenth ed.). Edinburgh: Oliver and Boyd.
- Friedman, D., & Sunder, S. (1994). *Experimental methods: a primer for economists*. Cambridge: Cambridge University Press.
- Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *Journal of Socio-Economics*, 40(11), 35–42.
- Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer & V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen, gwdg beric edn* (pp. 1–15). New York, NY: Worth Publishers.
- Huck, S., Normann, H. T., & Oechssler, J. (2004). Two are few and four are many: number effects in experimental oligopolies. *Journal of Economic Behavior & Organization*, 53(4), 435–446.
- Maniadis, Z., Tufano, F., & List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *American Economic Review*, 104(1), 277–290. doi:10.1257/aer.104.1.277.
- Roux C. Thöni C (2013) Collusion among many firms: the disciplinary power of targeted punishment. DEEP Working Paper 1302.
- Sitzia, S., & Zizzo, D. J. (2011). Does product complexity matter for competition in experimental retail markets? *Theory and Decision*, 70(1), 65–82.
- Sudgen, R., Zheng, J., & Zizzo, D. J. (2013). Not all anchors are created equal. *Journal of Economic Psychology*, 39, 21–31.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185(4157), 1124–1131.
- Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: basic anchoring and its antecedents. *Journal of Experimental Psychology*, 125, 387–402.

- Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1), 75–98.
- Zizzo, D. J., & Flemming, P. (2011). Can experimental measures of sensitivity to social pressure predict public good contribution? *Economics Letters*, 111(3), 239–242.