Electronic Companion to: On Repeat: Does Iteration Drive Innovation?

Evgeny Kagan
Johns Hopkins Carey Business School, Johns Hopkins University, ekagan@jhu.edu

Christian Jost
University of Augsburg, christian1.jost@uni-a.de

Sebastian Schiffels
University of Augsburg, sebastian.schiffels@uni-a.de

Tobias Lieberum
Unaffiliated, tobias.lieberum@tum.de

Abstract. This electronic companion is organized as follows. In EC.1 we present the details of the experimental procedures, sample sizes and exclusions, as well as the experimental interface, in particular the screenshots of each task. In EC.2 we present the parameter sets used in each task and treatment. In EC.3 we present several supplementary analyses that help support our results.

EC.1. Experimental Protocol and Exclusions

Experiments were programmed in oTree (Chen et al. 2016). Participants were recruited via ORSEE (Greiner et al. 2004). A total of 643 participants were recruited. A total of 43 participants were not admitted to the experiment because they had failed the German test. A total of 2 participants were excluded due to technical issues. A total of 89 participants were not admitted to the Scrabble task because they failed the attendant comprehension test, and a total of 43 participants were not admitted to the Lemonade Stand task because they failed the attendant comprehension test. If a participant was excluded from one of the two tasks, they did not receive the payoff from that task, and continued to the second task.

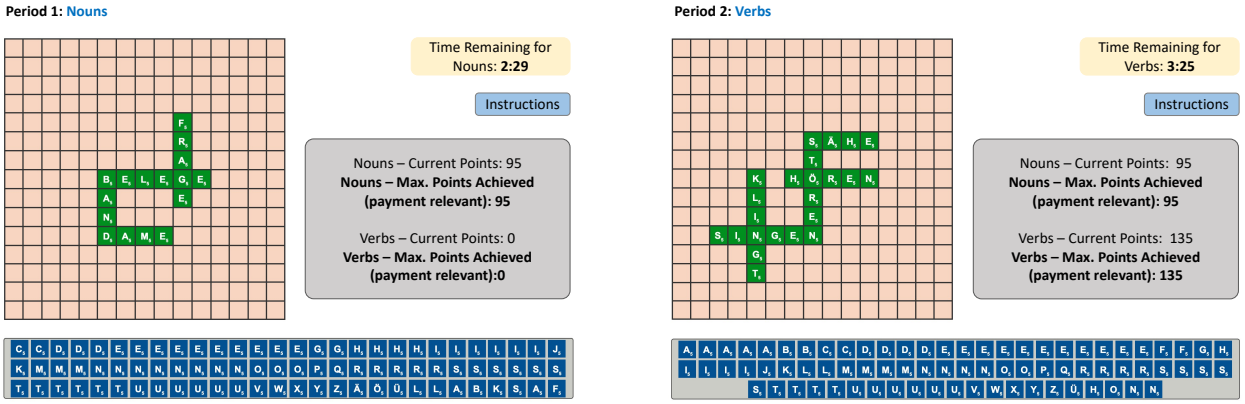
Due to Covid-19 restrictions during the initial wave of recruitment all experiments were conducted online. For consistency, the subsequent experimental waves were also conducted online even after the Covid-19 restrictions had been lifted. Zoom was used to monitor the participants. Zoom meetings were set up with at least one of the authors as a host. Participants received Zoom links via email in the morning of the day of the experiment. Upon sign-up, participants were renamed to preserve anonymity. During the experiment participants were able to chat with the experimenter and ask questions. All instructions were read aloud. To conserve space in this electronic compendium, the instructions were relegated to an online depository: `https://researchbox.org/3723&PEER_REVIEW_passcode=PZXHPR`.

EC.1.1. Scrabble Task (§4): Screenshots

Figure EC.1.1 shows the screenshots of the experimental interface for the *SEQ* treatment. The interface is analogous for the *ITER* treatment, with the addition of a button for switching between components.

Figure EC.1.1 Scrabble: Screenshots

(a) Task 1: Nouns (b) Task 2: Verbs



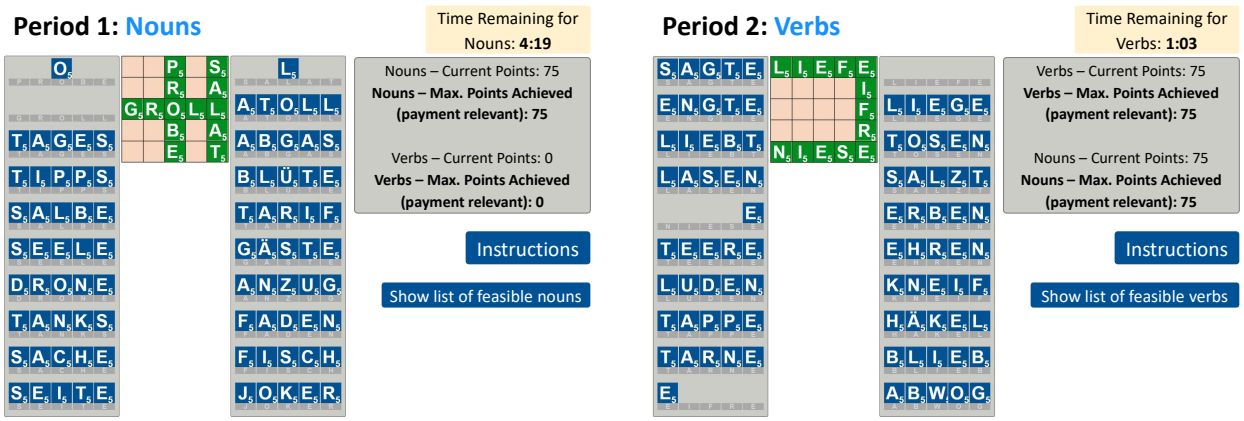
Note: Sample screenshots (translated from German) for the sequential workflow treatment.

EC.1.2. Scrabble with Pre-formed Words (§5): Screenshots

Figure EC.1.2 shows screenshots of the interface of the Scrabble with pre-formed words task (discussed in §5). Similar to the original Scrabble task, this task also consisted of two components: one board with nouns and a second board with verbs only. Each participant received the same list of 20 nouns and 20 verbs, displayed in Figure EC.1.2. As before each letter placed on the board was worth 5 points. There was a single "global " maximum (placing a total of six words on the board, resulting in a payoff of 150 points) and two suboptimal, "local" maxima (placing a total of five words, resulting in a payoff of 125 points). The remaining experimental details (task durations, random ordering etc.) were unchanged relative to the original Scrabble task of §4.

Figure EC.1.2 Scrabble with Pre-formed Words: Screenshots

(a) Task 1: Nouns (b) Task 2: Verbs

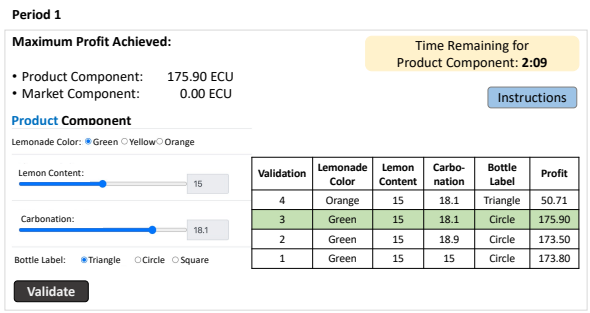


EC.1.3. Lemonade Task (§6): Screenshots

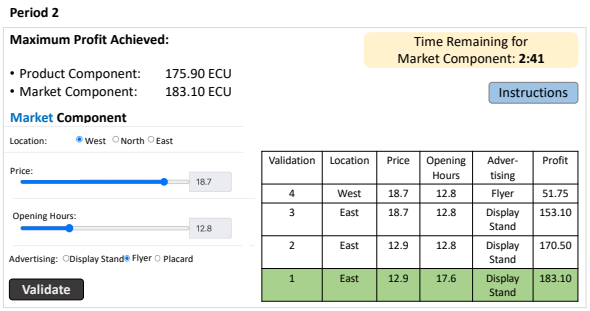
Figure EC.1.3 shows the screenshots of the experimental interface for the *SEQ* treatment. The interface is analogous for the *ITER* treatment, with the addition of a button for switching between components.

Figure EC.1.3 Lemonade Stand Task: Screenshots

(a) Component 1: Product



(b) Component 2: Market



Note: Sample screenshots (translated from German) for the *SEQ* treatment. As in previous work using the lemonade stand game (Ederer and Manso 2013, Sommer et al. 2020), we use a mix of discrete and continuous attributes. Specifically, lemonade color, bottle label, location and advertising are discrete attributes. The remaining attributes are continuous. The continuous attributes allow inputs in the $[10, 20]$ range, with the choices limited to one digit after the decimal point, yielding a total of 101 possible choices each. Thus, the solution space in each task has $3 \times 3 \times 101 \times 101 = 92,000$ unique combinations. The tables show each examined combination, with the best discovered combination highlighted in green.

EC.2. Experimental Design Details and Parametrization

In this section we describe the implementation details of all three tasks, in particular, the materials provided to participants during the Scrabble tasks, and the parametrization of the Lemonade Stand task.

EC.2.1. Scrabble Task (§4)

Following the classic German version of Scrabble, 100 tiles were made available to the subjects for each (Noun and Verb) task. The tiles were not refilled for the second period. The tiles given to participants at the beginning of the task were as follows (number of tiles with each letter is given in parentheses):

E (15), N (9), S (7), I (6), R (6), T (6), U (6), A (5), D (4) H (4), G (3), L (3), O (3) M (4), B (2), W (1), Z (1) C (2), F (2), K (2), P (1) Ä (1), J (1), Ü (1), V (1) Ö (1), X (1) Q (1), Y (1)

EC.2.2. Scrabble Task with Pre-formed Words (§5)

See Figure EC.1.2 for the complete list of words for this task.

EC.2.3. Lemonade Stand Task (§6)

We developed an adaptation of the classic Lemonade Stand game (Ederer and Manso 2013), which includes two components, each with a separate, independent solution landscape. The first component is the Product component, consisting of four attributes (lemonade color, lemon content, carbonation, shape of the bottle

label). The second component is the Market component, consisting of four attributes (location, price, opening hours, advertising). For each component, two of the attributes are discrete, with three levels to choose from, while the other two are chosen on the $\{10, 20\}$ range, and can be varied in increments of 0.1 units. Figure EC.2.1 and Figure EC.2.2 show the landscapes for all combinations of the discrete variables for the Product and the Market components, for the *Rugged* parametrization.

Subjects were presented with two components (Product and Market), with each component containing four parameters.

Product component:

1. Color = {Green, Yellow, Orange}
2. Lemon content = $\{10, 10.1, 10.2, ..., 19.9, 20\}$
3. Carbon dioxide content = $\{10, 10.1, 10.2, ..., 19.9, 20\}$
4. Bottle label = {Square, Triangle, Circle}

Market component:

Figure EC.2.1 Lemonade Stand Task: Market Component (Rugged Parametrization)

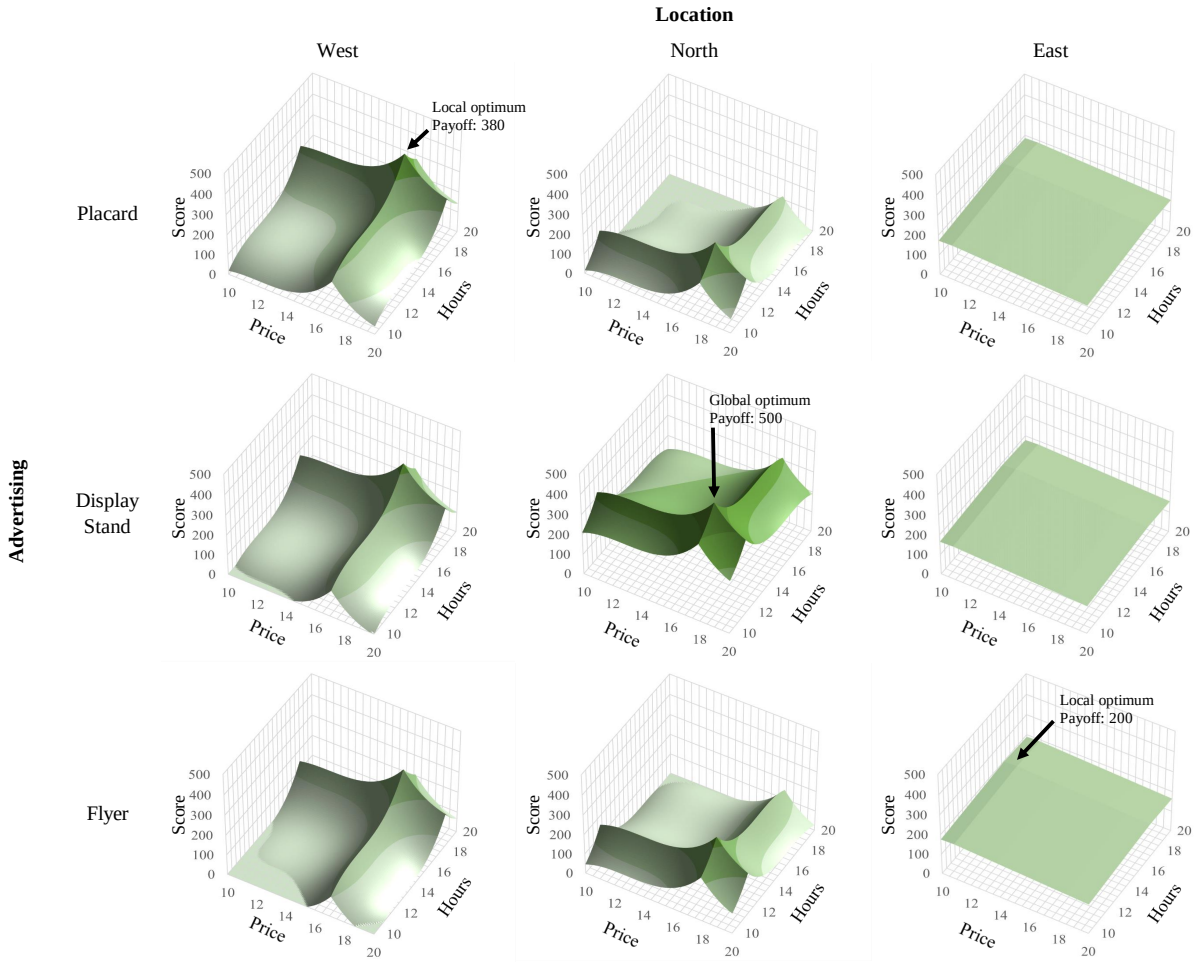
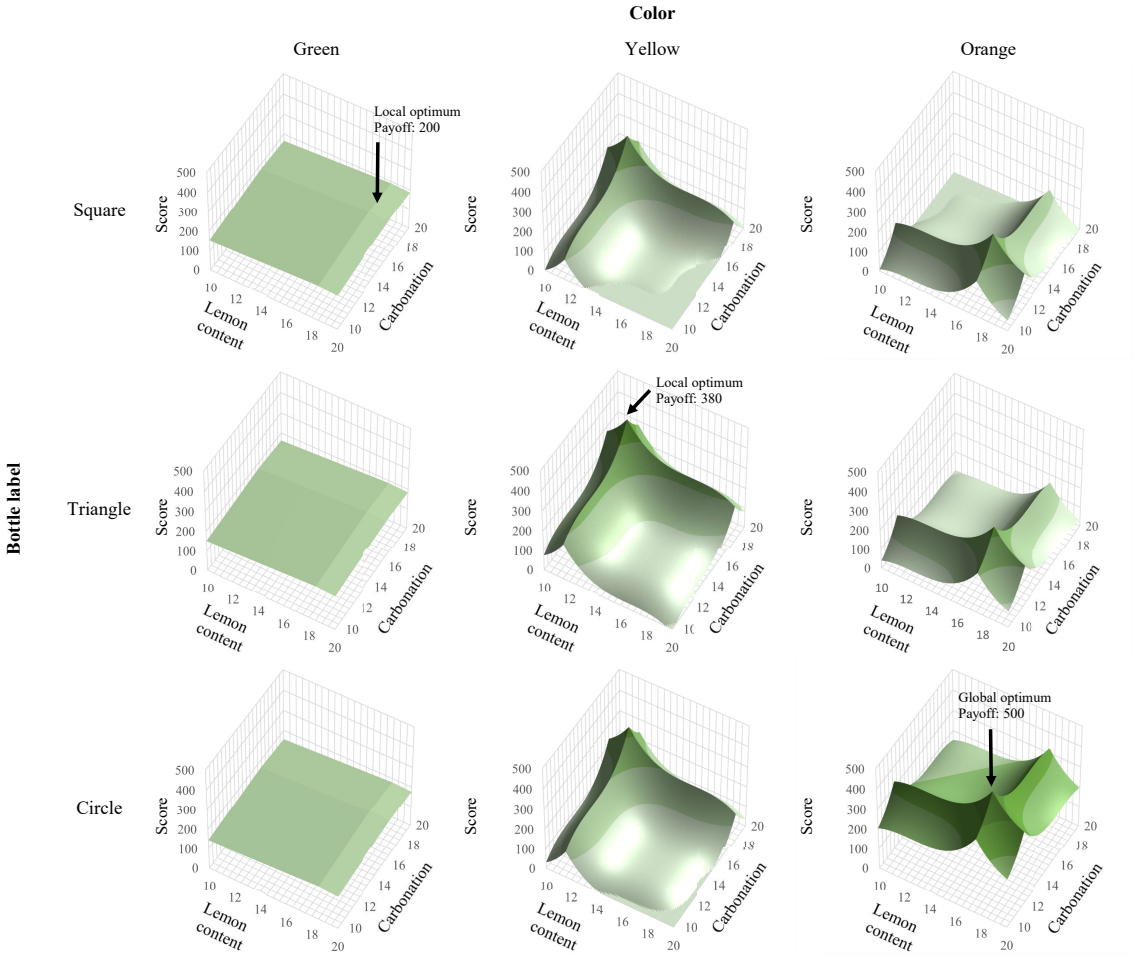


Figure EC.2.2 Lemonade Stand Task: Product Component (Rugged Parametrization)



1. Location = {West, North, East}
2. Price = $\{10, 10.1, 10.2, ..., 19.9, 20\}$
3. Opening hours = $\{10, 10.1, 10.2, ..., 19.9, 20\}$
4. Advertising = {Placard, Display Stand, Flyer}

EC.2.4. Rugged Parametrization (Treatments T1-T4)

We first described the parameters used in the more rugged parametrization of the Lemonade Stand game. For each lemonade color (in the Product component) and location (in the Market component), there is a predefined, optimal selection resulting in a maximum profit. To avoid the possibility that our effects were driven by a single parameter version we used two different parameter versions for each component. Table EC.2.1 shows the optimal selections and maximum profits for each component and version.

For the Market component, Figure EC.2.1 shows the three maxima, each of which corresponds to a combination of location and advertising. For the product component, Figure EC.2.2 shows the three maxima, each of which correspond to a combination of lemonade color and bottle label. As shown in Table

Table EC.2.1 Optimal Selection for *Rugged* Parametrization of Lemonade Stand Task

	Version 1			Version 2		
Product						
Lemonade color	Green	Yellow	Orange	Green	Yellow	Orange
Lemon content	18.5	11.6	17.6	11.5	12.4	18.4
Carbonation	16.9	18.5	12.2	13.1	17.8	11.5
Bottle label	Square	Triangle	Circle	Square	Triangle	Circle
Maximum Profit	200	380	500	380	500	200
Market						
Location	West	North	East	West	North	East
Price	17.1	17.9	10.9	12.9	19.1	12.1
Opening hours	18.5	11.8	17.3	11.5	12.7	18.2
Advertising	Placard	Display stand	Flyer	Placard	Display stand	Flyer
Maximum Profit	380	500	200	200	380	500

EC.2.1, we set these three maxima to 200, 380, and 500 points, respectively. Note that while the optimal locations of the remaining attributes are unchanged if we move vertically in Figures EC.2.1 and EC.2.2, the locations change if we move horizontally. This corresponds to the medium complexity scenarios used in the prior rugged landscape literature (see, for example, Sommer et al. 2020, and references there). The penalties for the discrete attributes (lemonade color and bottle label for the Product component, as well as Location and Advertising for the Market component) are given in Table EC.2.2. The penalties for the lowest local maximum (at 200) for the continuous attributes (Lemon content and Carbonation for the Product component, as well as Price and Opening hours for the Market component) are linear. They were computed by multiplying each unit of absolute deviation by a constant, i.e., $absolute\ deviation \times 3$. In order to achieve a sufficiently high level of difficulty, the penalty functions are S-shaped; that is, the gradient decreases the closer one gets to the optima. To achieve this, the penalty functions were calibrated as follows: $(\frac{absolute\ deviation}{5} - 1)^3 \times 150 + 150$. These penalty functions led to a level of difficulty that was found to be appropriate in pre-experimental pilots with 33 participants.

Table EC.2.2 Penalties by Component and Parameter Version

		Version 1			Version 2		
Product							
Lemonade color		Green	Yellow	Orange	Green	Yellow	Orange
Bottle label	Square	0	75	195	0	165	10
	Triangle	3	0	165	75	0	3
	Circle	10	45	0	45	195	0
Market							
Location		West	North	East	West	North	East
Advertising	Display stand	45	0	10	10	0	165
	Flyer	75	165	0	3	75	0
	Placard	0	195	3	0	45	195

EC.2.5. Smooth Parametrization (Treatments T5 and T6)

We next describe the parameters used in the smooth parametrization of the Lemonade Stand game (used in treatments T5 and T6, see Table 1 for treatment details). The parameters were chosen analogously to Table EC.2.1 and Table EC.2.2, with the difference that the optimal combination of the continuous variables in the Market component was always $Price = 18.2$ and $Hours = 11.5$, regardless of the remaining two attributes. Similarly, the optimal combination of the continuous variables in the Product component was always $Lemon\ content = 18.5$ and $Carbonation = 16.9$, regardless of the remaining two attributes. Figures EC.2.3 and EC.2.4 show the landscapes. Note that the local optima are located in the same position in all nine combinations of discrete attributes. This is analogous to the low complexity scenario in Sommer et al. (2020).

Figure EC.2.3 Lemonade Stand Task: Market Component (Smooth Parametrization)

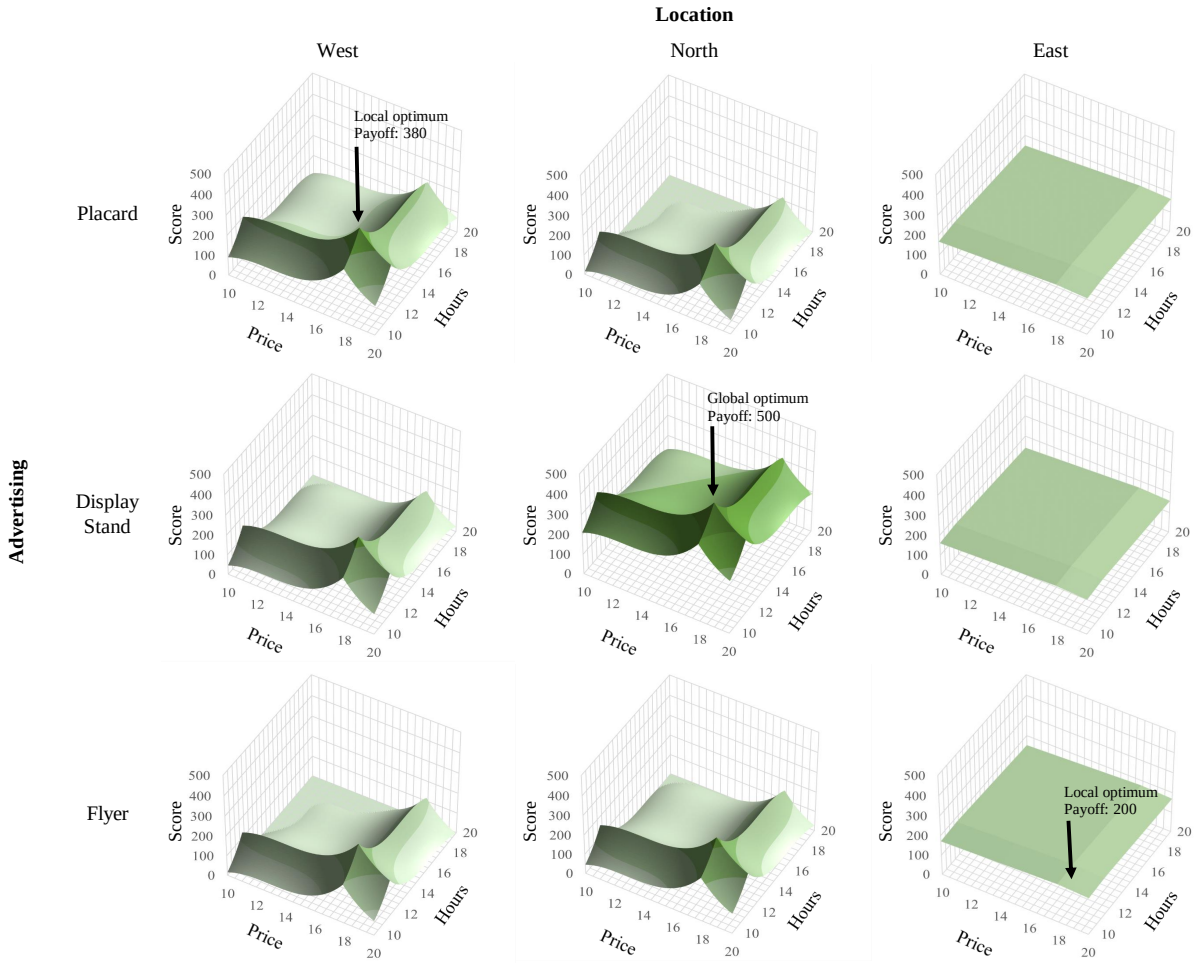
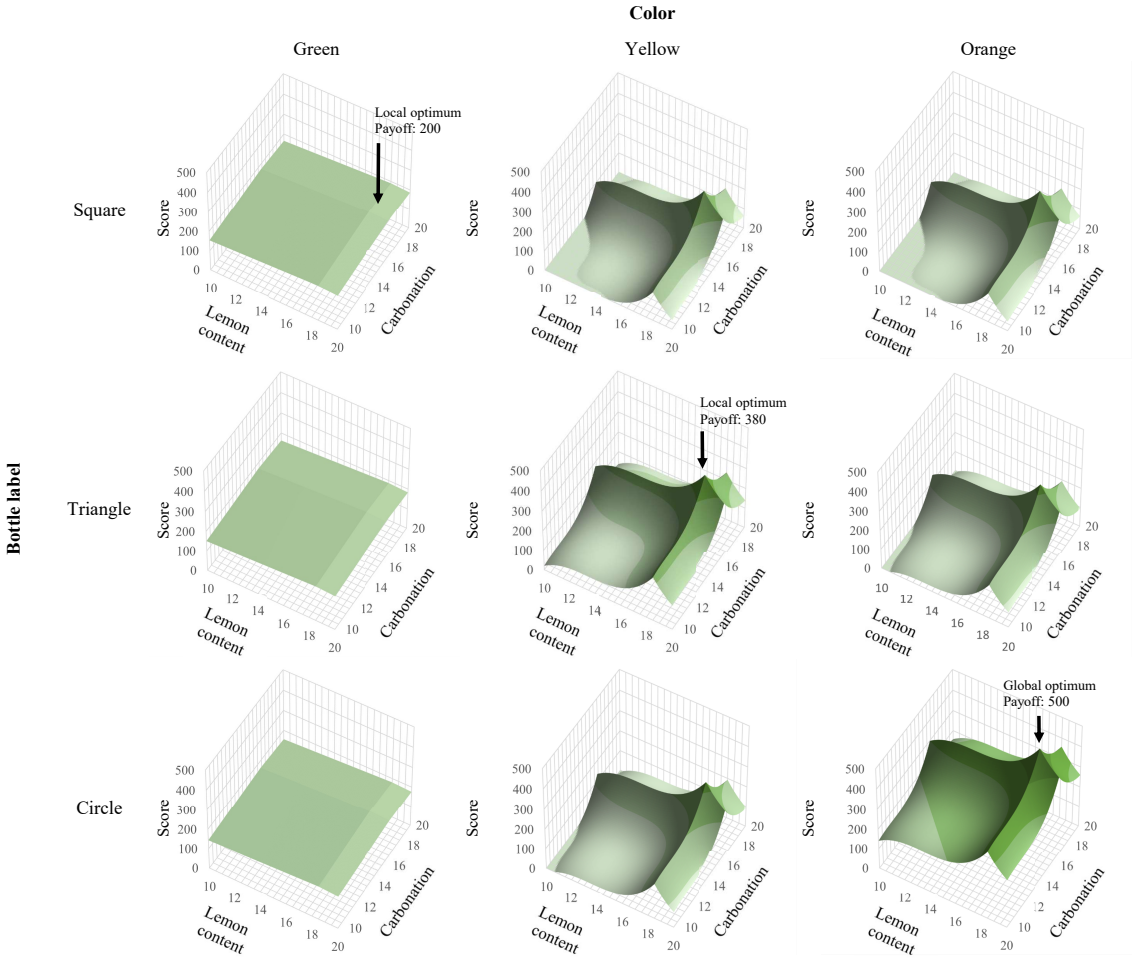


Figure EC.2.4 Lemonade Stand Task: Product Component (Smooth Parametrization)



EC.3. Additional Analysis

EC.3.1. Additional Analysis for §6: Treatments with Freezing

In the main text (§6) we discussed two treatments: *SEQ* and *ITER* with rugged and smooth parametrization. In addition to these treatments, we also conducted two treatments with freezing (in the rugged parametrization). We find that freezing significantly reduces performance in the Lemonade Stand task relative to both *SEQ* and *ITER* conditions. Figure EC.3.1 shows the performance in all six treatment conditions. Regression results are presented in Table EC.3.1 below. The percentage of participants reaching local and global optima is given in Table EC.3.2.

Recall that freezing in the Lemonade Stand task was achieved by fixing (during the second period) two of the four attributes of the landscape to their best discovered values. Our results show that this narrowing down of the solution landscape substantially reduced payoffs. The analysis in Table EC.3.2 shows that freezing substantially reduces the proportion of subjects reaching the global optimum region (from 19.6-24.0% to 9.6-9.9%).

Figure EC.3.1 Lemonade Stand Task: Performance by Treatment

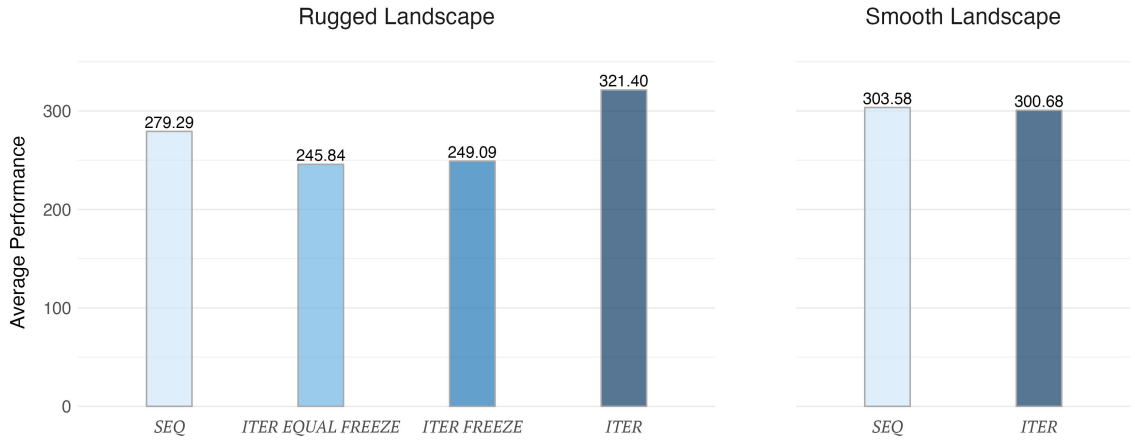


Table EC.3.1 Lemonade Stand Task: Regression Results

	Rugged landscape				Smooth landscape			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	*Performance*	*Performance*	*Period 1 Productivity*	*Period 2 Productivity*	*Performance*	*Performance*	*Period 1 Productivity*	*Period 2 Productivity*
SEQ	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline
ITER EQUAL FREEZE	-33.45**	-32.15**	214.7***	-274.4***				
	(14.86)	(15.13)	(21.18)	(17.09)				
ITER FREEZE	-30.20**	-29.70*	176.2***	-242.8***				
	(14.98)	(15.09)	(21.14)	(17.05)				
ITER	42.11**	44.36**	260.8***	-226.2***	-2.891	-5.448	176.7***	-205.6***
	(16.66)	(17.16)	(24.03)	(19.38)	(17.95)	(18.42)	(23.04)	(23.72)
Constant	279.3***	283.0***	292.0***	402.7***	303.6***	385.8***	317.6***	533.2***
	(9.886)	(36.76)	(51.48)	(41.54)	(12.42)	(52.68)	(65.90)	(67.84)
Dem. and session controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	286	286	286	286	115	115	115	115
R-squared	0.078	0.093	0.389	0.568	0.000	0.071	0.379	0.464

Notes. OLS regressions with standard errors in parentheses. Controls are age, gender, German native speaker, education, parameter version, task and component sequence. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table EC.3.2 Lemonade Stand Task: % of Subjects in Each Region

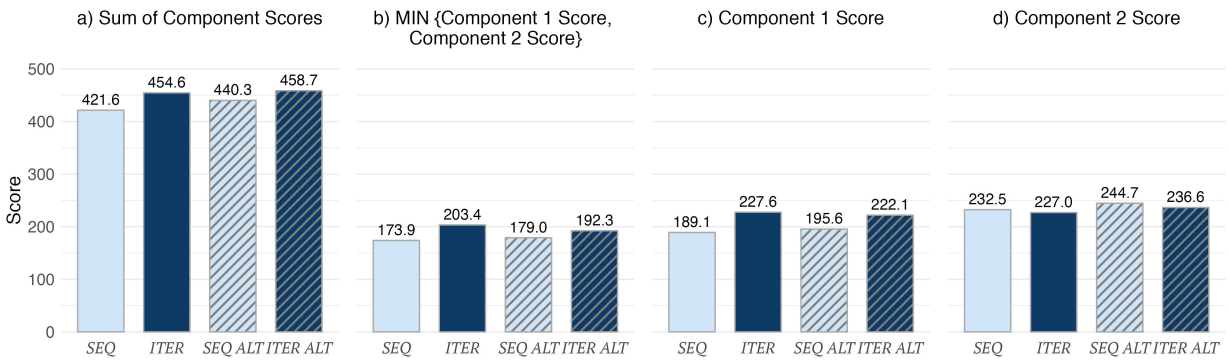
	Rugged Parametrization				**Smooth Parametrization**	
	SEQ	*ITER EQUAL FREEZE*	*ITER FREEZE*	*ITER*	*SEQ*	*ITER*
% of subjects reaching **global** optimum region	19.6%	9.6 %	9.9%	24.0%	18.3%	14.5%
% of subjects reaching **middle** local optimum region	28.3%	24.7%	28.2%	56.0%	46.7%	60.0%
% of subjects reaching **low** local optimum region	52.2%	65.8%	62.0%	20.0%	35.0%	25.5%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

EC.3.2. Additional Analysis for §7-8

EC.3.2.1. Scrabble Task: Summary Statistics Figure EC.3.2 shows performance in each treatment. Panels a) and b) show overall performance metrics, i.e., the sum of component scores and the minimum of component scores. First, the incentive scheme appears to have only minimal effects on performance: the differences between *SEQ* and *SEQ ALT* and between *ITER* and *ITER ALT* are relatively small. Second, the workflow effects are less pronounced if we compare overall performance (sum of component scores), than if we compare the minimum scores in panel b). In particular, in panel a) the differences between sequential and iterative treatments are directionally consistent with §4, but the magnitudes of the differences are smaller than in panel b).

Panels c)-d) help clarify the reasons for the overall performance results. In panel c) (resp. panel d) we examine Component 1 (resp.: Component 2) performance. Note that the first/second displayed component is interchangeable in the iterative workflow treatments, given that participants can immediately switch to the other component. The performance comparisons are as follows. First, in the sequential workflow, the second displayed component outperforms the first one, suggesting that most participants improve over time.[1] In contrast, in iterative workflow, performance differences between the first and the second components are minimal. As a result, if we focus on Component 1 performance, the differences between sequential and iterative regimes closely resemble the differences in the minimum scores shown in panel b). In particular, the treatment difference in Component 1 scores between sequential and iterative regimes is between 26.55 and 38.71 points, corresponding to a 20.35% to 12.5% gap. The initial slowdowns in sequential workflow therefore explain the more pronounced treatment differences in minimum-based performance.

Figure EC.3.2 Scrabble Task: Performance by Treatment



EC.3.2.2. Scrabble Task: Regression Analysis In this section we examine the interactions between workflow and incentive scheme. In Table EC.3.3 we focus on the Scrabble task. The results show that none of the interactions are statistically significant and the marginal effects of workflow are similar for different incentive schemes (bottom panel of Table EC.3.3).

[1] Specifically, the second component score is higher than the first one for 70% (resp.: 72%) of participants in *SEQ* (resp.: *SEQ ALT*).

Table EC.3.3 Scrabble Task: Analysis with Interaction Effects

Dependent Variable:	(1) *MIN{Component 1 Score, Component 2 Score}*	(2) *Sum of Component Scores*	(3) *Component 1 Score*	(4) *Component 2 Score*	(5) *Period 1 Productivity*	(6) *Period 2 Productivity*
SEQ	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline
ITER	23.48	20.69	31.07**	-10.38	44.01***	-23.31*
	(14.65)	(25.22)	(15.55)	(14.99)	(16.95)	(15.40)
Sum-based Incentive System (ALT)	8.09	30.10	10.46	19.64	10.07	20.03
	(14.61)	(25.16)	(15.51)	(14.95)	(16.90)	(15.36)
ITER × *ALT*	-7.99	1.64	-2.16	3.80	7.90	-6.25
	(19.46)	(33.49)	(20.65)	(19.91)	(22.51)	(20.45)
Constant	186.58***	414.18***	174.33***	239.86***	200.43***	213.76***
	(29.83)	(51.35)	(31.65)	(30.52)	(34.50)	(31.35)
Observations	276	276	276	276	276	276
R-squared	0.145	0.163	0.189	0.090	0.200	0.103
			p−values from post-hoc tests:			
ITER (Average effect)	0.047	0.190	0.004	0.398	0.000	0.008
ITER (at *ALT*= 0)	0.110	0.413	0.047	0.489	0.010	0.131
ITER (at *ALT*= 1)	0.223	0.307	0.033	0.613	0.000	0.027

Notes. OLS regressions with standard errors in parentheses. Columns (1)-(2) use overall performance, i.e., the minimum or the sum of the two component scores as the dependent variable. Columns (3)-(4) use period-wise productivity, i.e., the sum of score improvements achieved in each period. Columns (5)-(6) use component scores. All specifications control for task sequence, component sequence, and demographics. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

EC.3.2.3. Lemonade Stand Task: Summary Statistics Figure EC.3.3 shows total and component-wise scores by treatment. The results are similar to the Scrabble task. As before, iterative workflow outperforms sequential in Component 1. Further, somewhat differently from the Scrabble task, sequential treatments not only catch up, but outperform iterative treatments in Component 2. The effect is somewhat stronger with the alternative, sum-based incentive scheme. As a result, there appear to be minimal treatment differences in performance if we focus on the sum of component scores (panel a), and some differences in performance if we focus on the minimum of component scores (panel b). The slow learning in sequential regimes means that the minimum of the two scores is lower with sequential work, but the robust gains in the second period help them catch up when it comes to the sum of the two component scores.

EC.3.2.4. Lemonade Stand Task: Regression Analysis In Table EC.3.4 we focus on the Lemonade Stand task. The results show that none of the interactions are statistically significant (at $p < 0.05$) and the marginal effects of workflow are directionally similar for each incentive scheme (bottom panel of Table EC.3.4). However, some of the marginal effects are not significant (at $p < 0.05$) with the sum-based incentives.

EC.3.3. Task Ordering in Sequential Workflow

We supplement our analysis of task ordering with the regression results (§7.3). In Table EC.3.5, we replicate the specification in col. (1) in Tables EC.3.3 and EC.3.4, with the subsample of *SEQ* participants that

Figure EC.3.3 Lemonade Stand Task: Performance by Treatment

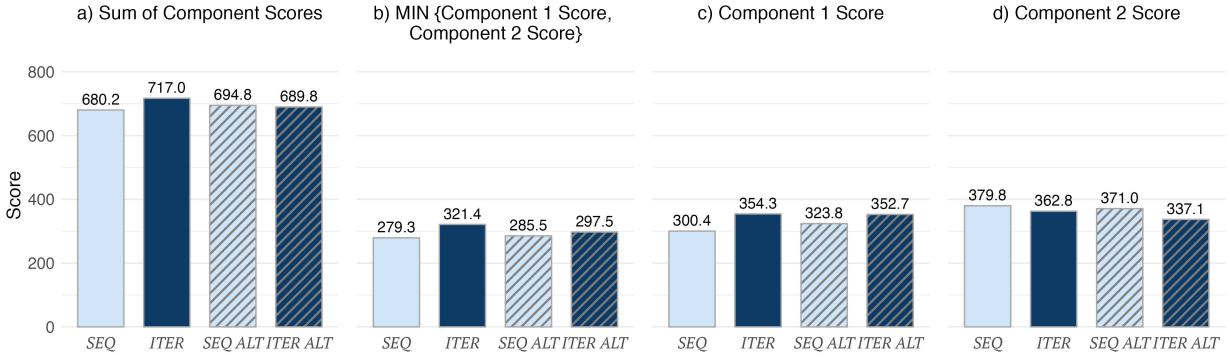


Table EC.3.4 Lemonade Stand: Analysis with Interaction Effects

Dependent Variable:	(1) *MIN{Component 1 Score, Component 2 Score}*	(2) *Sum of Scores*	(3) *Component 1 Score*	(4) *Component 2 Score*	(5) *Period 1 Productivity*	(6) *Period 2 Productivity*
SEQ	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline
ITER	49.57***	50.29	63.45***	-13.16	272.63***	-225.18***
	(18.64)	(31.61)	(20.07)	(20.43)	(25.42)	(22.11)
Sum-based Incentive System (ALT)	13.55	17.83	29.55	-11.72	29.12	-12.63
	(18.02)	(30.57)	(19.40)	(19.75)	(24.57)	(21.38)
ITER × *ALT*	-38.00	-50.30	-35.03	-15.26	-62.50*	13.44
	(25.00)	(42.39)	(26.91)	(27.39)	(34.08)	(29.65)
Constant	282.48***	733.68***	330.97***	402.71***	317.02***	424.27***
	(38.27)	(64.89)	(41.19)	(41.94)	(52.18)	(45.39)
Observations	296	296	296	296	296	296
R-squared	0.034	0.033	0.059	0.076	0.420	0.466
			p−values from post-hoc tests:			
ITER (Average effect)	0.018	0.257	0.001	0.125	0.000	0.000
ITER (at *ALT*= 0)	0.008	0.113	0.002	0.520	0.000	0.000
ITER (at *ALT*= 1)	0.490	1.000	0.116	0.123	0.000	0.000

Notes. OLS regressions with standard errors in parentheses. Columns (1)-(2) use overall performance, i.e., the minimum or the sum of the two component scores as the dependent variable. Columns (3)-(4) use period-wise productivity, i.e., the sum of score improvements achieved in each period. Columns (5)-(6) use component scores. All specifications control for task sequence, component sequence, and demographics. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

start with the weaker (Verbs in Scrabble, Product in Lemonade Stand) vs. start with the stronger (Nouns in Scrabble, Market in Lemonade Stand) component first. For simplicity, we do not include the interaction effects between workflow and incentive system (Doing so does not affect our results). The results suggest that the workflow effects are directionally correct in all subsamples. However, the effects are only statistically significant (at $p < 0.05$) if participants in sequential workflow start with the weaker component, i.e., Verbs in Scrabble task (col. 2 in Table EC.3.5) and Product in Lemonade Stand task (col. 4 in Table EC.3.5).

Table EC.3.5 Regression Results for *SEQ* subsamples

	Scrabble Task		Lemonade Stand (Rugged Landscape) Task	
	(1)	(2)	(3)	(4)
	Subsample: Nouns first	Subsample: Verbs first	Subsample: Market first	Subsample: Product first
Dep. Var.	$MIN\{$*Comp. 1 Score, Comp. 2 Score*$\}$		$MIN\{$*Comp. 1 Score, Comp. 2 Score*$\}$	
SEQ (subsample)	Baseline	Baseline	Baseline	Baseline
ITER	11.58	24.81**	18.88	30.39**
	(9.751)	(10.42)	(13.20)	(13.20)
Sum-based Incentive System (ALT)	-1.355	13.04	-13.32	5.666
	(11.92)	(12.80)	(15.58)	(15.73)
Constant	214.5***	187.8***	299.6***	270.9***
	(29.53)	(32.18)	(42.56)	(41.33)
Demographic controls	Yes	Yes	Yes	Yes
Observations	244	235	255	245
R-squared	0.152	0.128	0.027	0.028

Notes. OLS regressions with standard errors in parentheses. Demographic controls are age, gender, German native speaker, education. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

EC.3.4. Exploration Behavior

We supplement our analysis of exploration behavior (§7.4) with the following regression results.

In Table EC.3.6, we examine the differences in HHI during each attempt range. The analysis shows that the treatment differences are statistically significant throughout the task, with the iterative workflow participants engaging in broader landscape search.

In Table EC.3.7, we examine how exploration (measured using the Herfindahl-Hirschman Index (HHI)) moderates the effectiveness of iterative versus sequential task ordering on performance outcomes. Column

Table EC.3.6 Lemonade Stand: Exploration Behavior by Attempt Range

Dependent Variable:	(1)	(2)	(3)	(4)	(5)
Exploration $(1 - HHI)$ *during...*	*Attempts 1-5*	*Attempts 6-10*	*Attempts 11-15*	*Attempts 16-20*	*Attempts 21+*
SEQ	Baseline	Baseline	Baseline	Baseline	Baseline
ITER	0.0255**	0.0300**	0.0584***	0.0508***	0.0160
	(0.0106)	(0.0121)	(0.0132)	(0.0167)	(0.0216)
Sum-based Incentive System (ALT)	-0.0158	0.00559	-0.0185	0.00259	-0.0104
	(0.0125)	(0.0143)	(0.0154)	(0.0190)	(0.0237)
Constant	0.303***	0.284***	0.186***	0.188***	0.230**
	(0.0335)	(0.0382)	(0.0428)	(0.0529)	(0.0916)
Observations	295	295	288	260	180
R-squared	0.051	0.052	0.084	0.060	0.032

Notes. OLS regressions with standard errors in parentheses. All specifications control for position effects, parameter version, lab location, and demographics (age, gender, education). Dependent variable is the Herfindahl-Hirschman Index (HHI) achieved during each range of 5 attempts, with column (5) showing scores from attempts beyond the 20th attempt. Note that the number of observations is lower for later attempts. This is because the number of attempts differs by participant. For example, 35 participants make fewer than 20 attempts in each component, resulting in the number of observations dropping from 295 to 260 in col. 4. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table EC.3.7 Exploration and Iterative Workflow: Effect on Total Scores

Dependent Variable:	(1)	(2) *Product Component Score*	(3)
SEQ, Exploration $(1-HHI)$ *Quartile 1*	Baseline	Baseline	Baseline
ITER	31.35**	29.25**	61.17**
	(13.63)	(13.50)	(26.35)
Exploration $(1-HHI)$ *Quartile 2*		52.09***	56.50**
		(18.62)	(23.56)
Exploration $(1-HHI)$ *Quartile 3*		68.33***	83.18***
		(18.40)	(24.50)
Exploration $(1-HHI)$ *Quartile 4*		47.98**	83.10***
		(18.51)	(24.98)
Exploration $(1-HHI)$ *Quartile 2* $\times$ *ITER*			-2.34
			(37.67)
Exploration $(1-HHI)$ *Quartile 3* $\times$ *ITER*			-35.27
			(36.90)
Exploration $(1-HHI)$ *Quartile 4* $\times$ *ITER*			-76.30**
			(37.07)
Sum-based Incentive System (ALT)	5.70	9.59	10.15
	(16.04)	(15.83)	(15.77)
Constant	336.2***	293.2***	287.8***
	(43.03)	(45.43)	(44.58)
Observations	296	296	296
R-squared	0.040	0.064	0.107
			p−values from post-hoc tests:
Marginal effect of *ITER* at Exploration Quartile 1	–	–	0.021
Marginal effect of *ITER* at Exploration Quartile 2	–	–	0.033
Marginal effect of *ITER* at Exploration Quartile 3	–	–	0.324
Marginal effect of *ITER* at Exploration Quartile 4	–	–	0.561

Notes. OLS regressions with standard errors in parentheses. Marginal effects show the effect of iterative treatment at each quartile level. $^{*}p<0.10$; $^{**}p<0.05$; $^{***}p<0.01$.

(1) replicates the baseline effect of iterative workflow on performance, showing a significant positive coefficient of 31.35 points ($p<0.05$). Column (2) introduces exploration quartiles as main effects, and shows that greater exploration (quartiles 2-4) consistently improves performance relative to the lowest exploration quartile. Column (3) presents the full interaction model, showing that the benefits of iterative workflow depend on exploration behavior. Specifically, while iterative processes yield substantial performance gains for participants in the lowest exploration quartile (61.17 points, $p=0.021$), this advantage diminishes with increased exploration. The interaction term for quartile 4 is significantly negative (-76.30, $p<0.05$), indicating that highly exploratory participants actually perform worse under iterative compared to sequential ordering. Post-hoc tests confirm that the iterative treatment effect becomes statistically insignificant for par-

ticipants in quartiles 3 and 4 ($p > 0.10$), suggesting that the the benefits of iterative workflow are offset when exploration is high.

References

Chen DL, Schonger M, Wickens C (2016) oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* 9:88–97.

Ederer F, Manso G (2013) Is pay for performance detrimental to innovation? *Management Science* 59(7):1496–1513.

Greiner B, et al. (2004) The online recruitment system orsee 2.0-a guide for the organization of experiments in economics. Technical report.

Sommer SC, Bendoly E, Kavadias S (2020) How do you search for the best alternative? Experimental evidence on search strategies to solve complex problems. *Management Science* 66(3):1395–1420.