Online Appendix
A1. Tables and Figures

Variable	% Difference	T-test p-value
Number of monthly batches	0.995%	0.849
Average number of items per batch	0.788%	0.772
Average spending ($) per batch	0.499%	0.419
Average fraction of items: found	-0.107%	0.686
Average fraction of items: replaced	0.192%	0.917
Average fraction of items: refunded	1.593%	0.539
Average picking time (seconds)	-0.963%	0.518
Number of picked items	0.284%	0.632
Number of days of work	0.205%	0.560

Table A1 Randomization check. Comparison of shoppers in the treatment and control groups. The second column shows how much the value of a variable differs between the treated and control groups.

Model	(1)	(2)
Dependent variable	Refund rate	Picking time
Treatment × Task_Complexity_High	-0.1789**	-53.3483****
	(0.0893)	(9.6162)
Treatment	-0.1560**	-35.0867****
	(0.0661)	(4.2784)
Task_Complexity_High	0.0755	107.1844****
	(0.0733)	(8.0917)
Controls	Yes	Yes
Store FE	Yes	Yes
Date FE	Yes	Yes
Observations:	139,045	139,045
R^2	0.0467	0.6161

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A2 Heterogeneous treatment effect: the table shows how the batch complexity affects the impact of the 'wayfinding' technology on shoppers' performance.

Model Dependent variable	(1) Refund rate	(2) Picking time
Treatment × *Workload_High*	-0.2424**	-71.6252****
	(0.1064)	(7.9585)
Treatment	-0.1536***	-33.8468****
	(0.0587)	(5.0767)
Workload_High	0.1171	-84.1209****
	(0.0825)	(6.0133)
Controls	Yes	Yes
Store FE	Yes	Yes
Date FE	Yes	Yes
Observations:	139,045	139,045
R squared	0.0467	0.6161

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A3 Heterogeneous treatment effect: the table shows how the shoppers' workload affects the impact of the 'wayfinding' technology on their performance.

Model Dependent variable	(1) Refund rate	(2) Picking time
Treatment × *Experience* × *Task_Complexity_High*	-0.0003	-0.4352****
	(0.0009)	(0.0859)
Treatment × *Experience*	-0.0017***	-0.6607****
	(0.0006)	(0.0413)
Treatment × *Task_Complexity_High*	-0.1206	-1.3577
	(0.1303)	(14.7421)
Treatment	-0.0157	23.0144****
	(0.0920)	(5.9312)
Experience	-0.0029****	-1.4203****
	(0.0005)	(0.0308)
Task_Complexity_High	0.0583	219.4848****
	(0.1002)	(11.3927)
Experience × *Task_Complexity_High*	0.0000	-1.1224****
	(0.0006)	(0.0634)
Controls	Yes	Yes
Store FE	Yes	Yes
Date FE	Yes	Yes
Observations:	139,045	139,045
R^2	0.0468	0.6203

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A4 Heterogeneous treatment effect: the table shows how the past experience of shoppers and batch complexity affect the impact of the 'wayfinding' technology on their performance.

Dependent variable	Interquartile Difference	
	Refund rate	Picking time
Model	(1)	(2)
Treatment	-0.61****	-12.27
	(0.16)	(13.27)
Relative effect size	5.09%	0.79%
Store FE	Yes	Yes
Time FE	Yes	Yes
Observations:	18,561	18,561
R^2	0.11	0.16

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A5 The impact of the 'wayfinding' technology on the interquartile difference of the batch refund rate and batch picking time.

Model	(1)	(2)	(3)
Dependent variable	Work volume	# Picked items	Store exploration
Treatment	0.0076****	0.3016****	0.0543****
	(0.0021)	(0.0533)	(0.0116)
Relative effect size	3.16%	5.53%	32.48%
Date FE	Yes	Yes	Yes
Observations:	572,680	572,680	572,680
R^2	0.0080	0.0069	0.0010

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A6 The impact of the 'wayfinding' technology on the daily workload, number of picked items, and store exploration. This is a shopper-day level analysis where we control for date-level fixed effects.

Model	(1)	(2)	(3)
Dependent variable	Work volume	# Picked items	Store exploration
Treatment $\times$ *Experience*	0.0014****	0.0563****	0.0029****
	(0.0001)	(0.0042)	(0.0006)
Treatment	-0.0010	-0.1092***	0.0305****
	(0.0013)	(0.0332)	(0.0090)
Experience	0.0140****	0.3403****	0.0055****
	(0.0001)	(0.0029)	(0.0004)
Date FE	Yes	Yes	Yes
Observations:	572,680	572,680	572,680
R^2	0.5049	0.4704	0.0045

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A7 Heterogeneous treatment effect: the table shows how the past experience of shoppers affects the impact of the 'wayfinding' technology on the daily workload, number of picked items, and store exploration.

A2. The Importance of the Routing Feature: Suggestive Evidence

To show the importance of the routing feature in the 'wayfinding' technology, we next conduct supplementary analyses leveraging the data on actual activation of the routing feature by treated gig workers. Our goal is to provide suggestive evidence of whether the activation of this routing feature by treated gig workers drives the improvements in service quality and efficiency. Importantly, activating the routing tool requires shoppers to calibrate their mobile device at the beginning of a batch shopping session. This calibration process introduces a modest time and effort cost, which means that not all treated gig workers might choose to activate the tool for every batch. This natural variation in activation provides a valuable opportunity to examine the effects of actual routing/algorithmic tool engagement, rather than just treatment assignment. To this end, we estimate the following regression specification when only focusing on shoppers from the treatment group:

$$Y_{kij_t} = \beta_0 + \beta_1 Routing\_Activated_{kij_t} + \boldsymbol{\gamma} X_{ki} + \boldsymbol{\eta} Z_{kj_t} + D_t + S_k + \varepsilon_{kij_t}, \tag{A1}$$

where X_{ki} is the $Store\_familiarity_{ki}$, an indicator variable equal to 1 if store k is the most frequently visited store by shopper i in the pretreatment period. Z_{kj_t} denotes a vector of batch-specific characteristics for batch j fulfilled at time t in store k, including the number of delivery orders, the number of items to be picked, and the number of product categories. D_t and S_k represent date and shopper fixed effects, respectively. $Routing\_Activated_{kij_t}$ is a binary variable equal to 1 if the routing technology was activated by shopper i for that batch, and 0, otherwise. Note that this regression specification differs substantially from specification (1) presented in the main body of the manuscript. First, we include date and shopper fixed effects here, whereas specification (1) uses date and store fixed effects. Importantly, shopper fixed effects cannot be included in specification (1) due to multicollinearity with the *Treatment* variable. Second, because the current model includes shopper fixed effects, we exclude time-invariant shopper-level covariates such as *Experience* and *Gender*, which would otherwise be absorbed. Third, rather than estimating the effect of *Treatment* on various outcomes, this specification focuses solely on treated shoppers and estimates the effect of activating the routing feature.

Table A8 shows the results on how activating the routing feature impacts the quality and efficiency of service among gig workers. As it was mentioned above, this analysis is restricted to the treatment group shoppers who were assigned access to the wayfinding technology and thus includes shopper-level fixed effects to account for time-invariant individual heterogeneity. In estimating this specification, we first assume that the activation of the routing feature is conditionally exogenous, i.e., uncorrelated with unobserved time-varying factors after controlling for fixed effects and

Model	(1)	(2)
Dependent variable	Refund rate	Picking time
Routing Activated	-0.73****	-92.56****
	(0.08)	(4.88)
Controls	Yes	Yes
Shopper FE	Yes	Yes
Date FE	Yes	Yes
Observations:	69,614	69,614
R^2	0.14	0.77

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A8 The impact of activating the routing feature on the refund rate and batch picking time. We only focus on the shoppers in the treatment group and control for shopper and date-level fixed effects.

batch characteristics.[1] As shown in Table A8, activation of the routing feature is associated with statistically and economically significant improvements in both dimensions of the performance. Specifically, when the routing tool is activated, shoppers exhibit a lower refund rate, indicating an improvement in service quality and a reduction in item picking time, reflecting greater operational efficiency. These effects underscore the potentially important role of the routing functionality in terms of how wayfinding technology influences outcomes. Notably, the second column of Table A8 shows that activating the routing feature reduces picking time by approximately 92 seconds per batch. These findings suggest that the algorithmic routing support might not be merely a peripheral feature, but might be an important component driving the observed treatment effects reported throughout the paper.

While the previous analyses assumed that the decision to activate the routing feature is exogenous, in reality, this assumption may not hold. For instance, shoppers may be more likely to activate the routing functionality when they feel less confident about the location of items or are less familiar with the store. If such behavior drives the endogeneity, then the estimates in Table A8 likely understate the true effect of the routing feature on key performance outcomes such as refund rate and picking time. To address this potential endogeneity in routing activation, we estimate the following instrumental variable (IV) regression:

$$Y_{kij_t} = \beta_0 + \beta_1 \widehat{Routing\_Activated}_{kij_t} + \boldsymbol{\gamma} X_{ki} + \boldsymbol{\eta} Z_{kj_t} + D_t + S_k + \varepsilon_{kij_t}, \quad \text{(A2)}$$

$$\widehat{Routing\_Activated}_{kij_t} = \delta_0 + \delta_1 Lagged\_Glitch_{kij_t} + \boldsymbol{\gamma} X_{ki} + \boldsymbol{\eta} Z_{kj_t} + D_t + S_k + \varepsilon_{kij_t}, \quad \text{(A3)}$$

where $Lagged\_Glitch_{kij_t}$ is a binary indicator equal to 1 if shopper i, at store k, experienced a calibration glitch with the routing feature during the previous batch completed before batch j at time t. The idea is that such glitches, driven by device connectivity or synchronization issues,

[1] We relax this assumption in our analyses below.

Model	(1)	(2)
Dependent variable	Refund rate	Picking time
Routing Activated	-1.13****	-137.40****
	(0.11)	(8.54)
Controls	Yes	Yes
Shopper FE	Yes	Yes
Date FE	Yes	Yes
Observations:	66,587	66,587
R^2	0.002	0.66

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A9 IV regression. The impact of activating the routing feature on the refund rate and batch picking time. We only focus on the shoppers in the treatment group and control for shopper and date-level fixed effects.

temporarily prevent the shopper from using the routing tool and might be exogenous to current batch/store conditions or shopper intent. The IV regression results are reported in Table A9. Consistent with our expectations, we continue to observe significant reductions in both refund rate and picking time when the routing feature is activated, even after instrumenting for its usage. More specifically, the second column in Table A9 indicates that activation of the routing feature leads to a 137-second reduction in the picking time. These findings provide additional suggestive evidence that the routing functionality might be an important feature of the wayfinding technology. Moreover, when comparing the IV estimates in Table A9 with the corresponding OLS results in Table A8, we find that the magnitudes of the estimated effects are notably larger under the IV specification. This pattern is consistent with our prior belief that OLS estimates may be biased due to negative selection, i.e., shoppers are more likely to activate the routing tool in more challenging or unfamiliar conditions, thereby underestimating the effect sizes.

We also present the first-stage regression results from equation (A3) in Table A10. The results confirm a strong and statistically significant negative relationship between the occurrence of a prior calibration glitch and the likelihood of routing tool activation in the current batch. The F-statistic in the first stage is 50.72, well above the conventional threshold of 10, confirming the strength of the instrument. Importantly, the validity of our instrument relies on the assumption that these calibration glitches are driven by random, system-level connectivity or synchronization problems that are not predictable by the shopper and are orthogonal to unobserved components in the context of current batch performance. While our IV regression helps mitigate this specific endogeneity concern, glitches may also capture broader device or connectivity problems that directly affect performance, so the exclusion restriction may not be fully satisfied. Accordingly, the IV estimates in this section should be interpreted as suggestive and potentially upward-biased.

Model	(1)
Dependent variable	Routing Activated
Lagged Calibration Glitch	-0.017****
	(0.004)
Controls	Yes
Shopper FE	Yes
Date FE	Yes
Observations:	66,587
R^2	0.233

** $p<0.05$, *** $p<0.01$, **** $p<0.001$

Table A10 IV regression: first stage results. We only focus on the shoppers in the treatment group and control for shopper and date-level fixed effects.

Next, one of the central findings in this paper is that experience plays a vital moderating role in the success of human–algorithm collaboration. We observe that the impact of the wayfinding technology is particularly strong for more experienced shoppers. However, an alternative interpretation of this result is that experienced shoppers may be more familiar with store layouts and therefore better able to interpret the visual ML-based map shown to them, independent of any algorithmic routing support. If true, this would suggest that the observed benefits for more experienced shoppers stem more from the ML-enabled map than from interacting with the routing algorithm itself. To provide suggestive evidence that the routing feature might also drive the interaction between experience and performance, we estimate a modified version of regression specification (A1), adding an interaction term between *Routing_Activated* and *Experience*. As defined in the main body of the manuscript, *Experience* is the number of batches fulfilled by a shopper at a given store during the pretreatment period of time. The results of this regression are reported in Table A11, which highlights the moderating effect of experience on the relationship between routing activation and performance outcomes. In particular, we find that activating the routing feature might lead to larger reductions in both refund rate and picking time for more experienced shoppers compared to less experienced ones. This finding suggests that the routing feature of the decision support tool is not only effective on its own but might be especially beneficial for more experienced shoppers, which is consistent with findings outlined earlier in the manuscript. Note that we do not include a separate term *Experience* in the regression specification in order to avoid multicollinearity, as shopper fixed effects are already included.

We then replicate the aforementioned analysis using an instrumental variables approach. In Table A12, we present the results of an IV specification in which $Lagged\_Glitch_{kijt}$ serves as an instrument for the potentially endogenous variable $Routing\_Activated_{kijt}$, and the interaction term $Lagged\_Glitch_{kijt} \times Experience_{ki}$ is used as an instrument for $Routing\_Activated_{kijt} \times$

Model	(1)	(2)
Dependent variable	Refund rate	Picking time
Routing Activated $\times$ *Experience*	-0.0017**	-0.15***
	(0.0008)	(0.05)
Routing Activated	-0.5751****	-78.30****
	(0.1109)	(6.86)
Controls	Yes	Yes
Shopper FE	Yes	Yes
Date FE	Yes	Yes
Observations:	69,614	69,614
R^2	0.1447	0.77

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A11 Moderating role of experience. We only focus on the shoppers in the treatment group and control for shopper and date-level fixed effects.

Model	(1)	(2)
Dependent variable	Refund rate	Picking time
Routing Activated $\times$ *Experience*	-0.0049****	-0.43****
	(0.0013)	(0.11)
Routing Activated	-0.2750**	-52.04****
	(0.1380)	(11.53)
Controls	Yes	Yes
Shopper FE	Yes	Yes
Date FE	Yes	Yes
Observations:	66,587	66,587
R^2	0.0027	0.66

Table A12 IV regression. Moderating role of experience. We only focus on the shoppers in the treatment group and control for shopper and date-level fixed effects.

$Experience_{ki}$. The results are consistent with our earlier estimates and provide suggestive evidence that the routing feature might play an important role in shaping the observed complementarities between experience and performance. Together, these results provide some partial evidence for the view that the routing feature of the decision support tool can especially amplify the productivity of experienced gig workers, underscoring the importance of aligning technology design with worker experience in human–algorithm collaboration contexts. As discussed above, this IV approach alleviates the endogeneity concern only partially, and thus, the exclusion restriction may not be fully satisfied. Therefore, these IV estimates should be viewed as suggestive rather than fully causal.

A3. Moderating Role of Experience: Model Free Evidence and Robustness

In this section, we first show the model-free evidence of the moderating role of experience in human-algorithm collaboration. To this end, we divide shoppers into five groups based on their experience level and examine how the treatment (i.e., access to the ML and algorithm-enabled decision support tool) affects their performance in terms of picking time and refund rate. These results are visualized in Figure A1. As shown, performance improvements generally increase with experience: more experienced shoppers benefit more from the technology, both in reducing refund rates and improving picking times. Notably, shoppers in the lowest experience group (first bin) exhibit only a marginal reduction in refund rates (much less than 1%) and actually experience an increase in picking time. This pattern reinforces our earlier finding that less experienced shoppers may not benefit from the real-time task support tool and can even perform worse when using it. This model-free evidence is consistent with the main results presented in the main body of the paper.

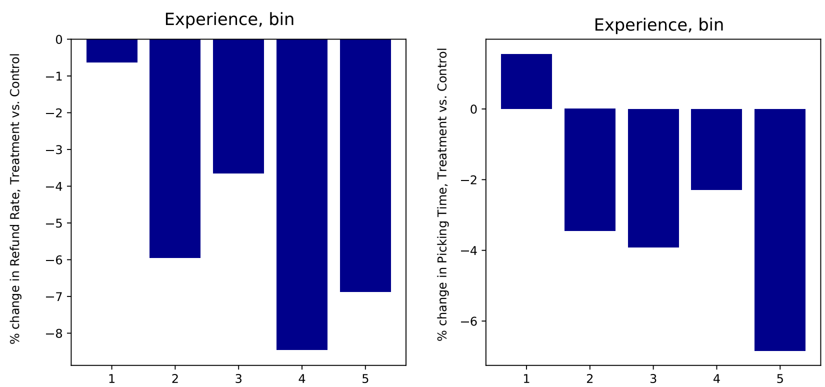

Figure A1 Each bar in those two plots corresponds to a particular subset of shoppers with a particular experience level; the higher the bin number the higher is the experience level. The left (resp. right) plot represents the relationship between shoppers' experience and the % change in the refund rate (resp. picking time) of the treatment shoppers versus control shoppers for a particular subset of shoppers corresponding to a particular experience level.

Finally, we demonstrate that our qualitative findings regarding the moderating role of experience remain robust even when experience is measured as a binary variable using a median split. Recall that, in the main body of our manuscript, experience corresponds to the number of batches a shopper fulfilled in a given store during the pretreatment period. Table A13 provides the robustness results and shows that our main conclusions hold. More experienced shoppers continue to benefit more from the technology, both in terms of improved service quality (as measured by a lower refund rate) and enhanced service efficiency (reflected in shorter picking times). In contrast, the effect of the technology on the less experienced shoppers appears to be either insignificant or negative. Specifically, less experienced shoppers do not see an improvement in refund rate and experience longer picking times when using the technology.

Model Dependent variable	(1) Refund rate	(2) Picking time
Treatment $\times$ *Experience High*	-0.2107***	-157.1003****
	(0.0809)	(6.6593)
Treatment	-0.1137	23.3663****
	(0.0688)	(5.7399)
Controls	Yes	Yes
Store FE	Yes	Yes
Date FE	Yes	Yes
Observations:	139,045	139,045
R^2	0.0465	0.6139

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A13 Heterogeneous treatment effect: the table shows how the higher experience of shoppers affects human-algorithm collaboration. The experience is the binary variable obtained using a median split.

A4. Task Complexity Balance

In order to address the concern regarding potential confounding between treatment assignment and task complexity, particularly given the dynamic nature of the ML algorithm that may, in theory, assign more complex tasks to more skilled workers, we empirically test whether task complexity is balanced across the treatment and control groups. Specifically, we examine whether shoppers in the treatment group systematically received batches with higher complexity than those in the control group. Consistent with the definition used in the main body of the paper, we measure task complexity as the number of product categories in a batch. We then construct a binary outcome variable, *Task Complexity High*, which equals 1 if the number of product categories exceeds the median and 0, otherwise. To this end, we estimate the balance of task complexity between treatment and control group shoppers using the following regression specification:

$$Y_{kij_t} = \beta_0 + \beta_1 Treatment_i + \boldsymbol{\gamma} X_{ki} + \boldsymbol{\eta} Z_{kj_t} + D_t + S_k + \varepsilon_{kij_t}, \tag{A4}$$

where Y_{kij_t} is the *Task Comlexity High*, which is a binary indicator set to 1 if the number of product categories in a batch exceeds the median and 0, otherwise, for the batch j_t fulfilled at time t at store k by shopper i, X_{ki} is the vector of the individual specific variables (where k corresponds to a particular store and i corresponds to a particular shopper) such as *Store familiarity*, *Experience*, and *Gender*. $Store\ familiarity_{ki}$ is an indicator variable and is equal to 1 if the store k is the store most visited by a shopper i based on the pretreatment data. $Experience_{ki}$ is the number of batches fulfilled by the shopper i of the store k based on the pretreatment data. Z_{kj_t} is the vector of batch-specific variables (where j_t corresponds to a particular batch fulfilled at time t) such as the number of delivery orders, the number of items to be picked and the number of product categories in a batch j fulfilled in the store k at time t. Note that every batch usually consists of around one to

Model	(1)
Dependent variable	Task Complexity High
Treatment	0.0031
	(0.0020)
Controls	Yes
Store FE	Yes
Date FE	Yes
Observations:	139,045
R^2	0.5070

$^{**}\ p < 0.05$, $^{***}\ p < 0.01$, $^{****}\ p < 0.001$

Table A14 Treatment does not effect the assignment of complex batches.

three delivery orders. D_t and S_k represent fixed effects at the date and the store level, respectively. Results from this regression are presented in Table A14. The coefficient on the treatment indicator is statistically insignificant, indicating that task complexity was balanced across treatment and control groups. This supports the validity of our main findings regarding heterogeneity by task complexity, confirming that they are not confounded by endogenous task assignment based on shopper performance.

A5. Robustness: Excluding Shoppers with Very Few Observations

In this section, we revisit the findings on how the complementarity between experience and algorithm-enabled technology varies by workers' skill levels. As shown in the main body of the manuscript, this human-algorithm collaboration appears particularly beneficial for lower-skilled workers, who gain more from the combination of accumulated experience and algorithmic support. This suggests that human-algorithm collaboration is most effective when workers have greater potential for improvement. To address concerns about potential measurement noise in defining skill for very inexperienced shoppers who have very few observations during the pretreatment period, we validate these findings by restricting the sample to those with a sufficient number of observations in the pretreatment period. Specifically, we reestimate our models using only shoppers who completed at least 10 and 25 batches during the pretreatment period. The results, reported in Tables A15 and A16, confirm that our insights remain qualitatively the same under the updated setup. It reinforces our conclusion that ML and algorithm-enabled technology most effectively complements experience among lower-skilled workers.

A6. Robustness: Human-Algorithm Collaboration

To address the concern that inexperienced workers might be disproportionately shopping during busier times of day or at more crowded store locations, potentially confounding the observed heterogeneity in treatment effects, we conduct additional robustness checks. Specifically, we examine

Model	(1)	(2)
Dependent variable	Refund rate	Picking time
Treatment × *Experience* × *Lower_Skilled*	-0.0025**	-0.4527****
	(0.0011)	(0.0902)
Treatment × *Experience*	-0.0003	-0.7250****
	(0.0007)	(0.0553)
Treatment × *Lower_Skilled*	-0.0041	-2.8437
	(0.1656)	(13.1990)
Treatment	-0.1229	52.0312****
	(0.1080)	(9.1536)
Experience	-0.0024****	-0.9674****
	(0.0005)	(0.0420)
Lower_Skilled	1.7902****	655.6660****
	(0.1210)	(11.9478)
Experience × *Lower_Skilled*	0.0008	-0.4607****
	(0.0008)	(0.0609)
Controls	Yes	Yes
Store FE	Yes	Yes
Date FE	Yes	Yes
Observations:	116,046	116,046
R^2	0.0606	0.6470

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A15 Heterogeneous treatment effect: the table shows how the past experience and performance of shoppers affect the impact of the 'wayfinding' technology on their performance. Herein, we only include shoppers who fulfilled at least 10 batches before the start of the experiment.

whether our main findings about the moderating role of experience hold when restricting the sample to (i) batches fulfilled during the first half of the day (6 a.m. to 3 p.m.), when store congestion is likely lower, and (ii) batches completed at high-popularity stores, based on a median split of store-level batch volumes. These sample restrictions help ensure that treatment and control shoppers perform their tasks under comparable working conditions. Results reported in Tables A17 and A18 confirm that our findings remain qualitatively the same in these subsamples. These robustness checks help alleviate concerns that time-of-day effects or store-specific congestion patterns are driving the observed differences in how experience moderates the impact of the real-time decision support tool.

A7. Robustness to Log Transformation

In this section, we show that our main finding – that the decision support technology helps lower-skilled workers narrow the performance gap with higher-skilled colleagues – remains robust when we log-transform the outcome variables (see Table A19). The log specification also addresses the alternative explanation that lower-skilled workers improve more simply because they have greater room to improve. Using multiplicative models based on log-transformed dependent variables, which measure relative effects, we find that the sign and statistical significance of the *Treatment* ×

Model	(1)	(2)
Dependent variable	Refund rate	Picking time
Treatment × *Experience* × *Lower_Skilled*	-0.0035***	-0.5766****
	(0.0012)	(0.1001)
Treatment × *Experience*	-0.0007	-0.7006****
	(0.0008)	(0.0601)
Treatment × *Lower_Skilled*	0.1749	52.0785***
	(0.1987)	(15.9079)
Treatment	-0.0548	34.8842***
	(0.1331)	(10.5819)
Experience	-0.0021****	-0.9992****
	(0.0006)	(0.0467)
Lower_Skilled	1.6758****	593.2337****
	(0.1481)	(14.2812)
Experience × *Lower_Skilled*	0.0014	-0.1908***
	(0.0009)	(0.0680)
Controls	Yes	Yes
Store FE	Yes	Yes
Date FE	Yes	Yes
Observations:	97,328	97,328
R^2	0.0640	0.6527

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A16 Heterogeneous treatment effect: the table shows how the past experience and performance of shoppers affect the impact of the 'wayfinding' technology on their performance. Herein, we only include shoppers who fulfilled at least 25 batches before the start of the experiment.

Model	(1)	(2)
Dependent variable	Refund rate	Picking time
Treatment × *Experience*	-0.0021****	-0.8915****
	(0.0006)	(0.0574)
Treatment	-0.0241	39.6460****
	(0.0932)	(8.8175)
Experience	-0.0016***	-1.8470****
	(0.0005)	(0.0438)
Controls	Yes	Yes
Store FE	Yes	Yes
Date FE	Yes	Yes
Observations:	72,565	72,565
R^2	0.0566	0.6161

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A17 Heterogeneous treatment effect: the table shows how the past picking experience of shoppers affects the impact of the 'wayfinding' technology on their performance. We only focus on batches fulfilled during the first half of the day.

Lower_Skilled interactions are unchanged. This reinforces our interpretation that the technology is especially beneficial for lower-skilled workers.

Model	(1)	(2)
Dependent variable	Refund rate	Picking time
Treatment × *Experience*	-0.0032****	-0.9909****
	(0.0005)	(0.0443)
Treatment	0.0972	38.3349****
	(0.0836)	(7.0597)
Experience	-0.0024****	-1.7648****
	(0.0004)	(0.0319)
Controls	Yes	Yes
Store FE	Yes	Yes
Date FE	Yes	Yes
Observations:	109,595	109,596
R^2	0.0517	0.6156

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A18 Heterogeneous treatment effect: the table shows how the past picking experience of shoppers affects the impact of the 'wayfinding' technology on their performance. We only focus on batches fulfilled at high-popularity stores.

Model	(1)	(2)
Dependent variable	Refund rate	Picking time
Treatment × *Lower_Skilled*	-0.0250**	-0.0131**
	(0.0124)	(0.0061)
Treatment	-0.0295****	-0.0200****
	(0.0086)	(0.0042)
Lower_Skilled	0.2650****	0.3514****
	(0.0088)	(0.0062)
Controls	Yes	Yes
Store FE	Yes	Yes
Date FE	Yes	Yes
Observations:	139,045	139,045
R^2	0.1359	0.6259

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table A19 Heterogeneous treatment effect: the table shows how the past performance of shoppers affects the impact of the 'wayfinding' technology on their current performance. The outcome variables in this table are log-transformed.