SHORT- AND LONG-TERM EFFECTS OF NONCONSCIOUSLY PROCESSED AMBIENT SCENTS IN A SERVICESCAPE: FINDINGS FROM TWO FIELD EXPERIMENTS

ONLINE APPENDIX

Online Appendix A: Pretest on the Nonconscious Processing of the Ambient Scent

To determine the olfactory stimulus's optimal intensity level for the two field experiments, we conducted a series of field pretests (n = 198 in scented train compartments and n = 132 consumers as control group without ambient scent) with actual consumers of the railway company on a specific rail section between two midsize towns in Germany. We surveyed n = 198 consumers (57% female, between 15 and 64 years old) with three different scent intensities (low, medium, and high, between-subjects, one intensity per consumer), dependent on the number of scent cartridges used per coach (i.e., four, six, or eight cartridges). For the n = 132 consumers in the control group (54% female, between 16 and 77 years old), no such manipulation was applied. The experiments took place in the same train, on the same track section, with similar consumers, and comparable temperatures, noise level, etc.

In each of the intensity levels, we first surveyed whether the consumers noted any special scent (i.e., unaided). We subsequently made all the consumers aware of the scent diffusion and again surveyed their perception (i.e., aided) (Doucé et al. 2013). In the lowest intensity level (4 cartridges), 25 of 86 participants (29%) consciously perceived the olfactory stimulus (11 unaided, 14 aided). When using six cartridges, 15 of 50 (30%) perceived the scent (8 unaided, 7 aided),

and even with the highest intensity level of eight scent cartridges per coach, only 13 of 62 (21%) were able to perceive the scent consciously (5 unaided, 8 aided).

We found no significant relationship between scent recognition and scent intensity ($\chi^2_{(2)} = 1.563$, p = .458, Cramer's V = .089). With an overall detection rate of 27% (53 participants who perceived the scent consciously), our results clearly indicate a nonconscious processing for the majority of consumers, which corresponds exactly to the rate observed in related research (Krishna, Lwin, and Morrin 2010).

As mentioned in the main article, we proceeded with further analyses involving the $n = 53$ consumers who noticed the scent stimulus. In addition to the information provided in the main article, these consumers' mean ratings of perceived intensity on a seven-point bipolar scale (-3 = very weak / +3 = very strong; Spangenberg, Crowley, and Henderson 1996) did not differ significantly between the three intensity levels ($M_{low} = -.26$, $SD_{low} = 1.54$; $M_{medium} = .73$, $SD_{medium} = 1.53$; $M_{high} = .25$, $SD_{high} = 2.01$; ANOVA: $F_{(2,49)} = 1.601$, p = .213, $\eta^2 = .064$).

As we found no significant differences in the perceived intensity between the three intensity levels, we collapsed across the three scent conditions for further analyses. We examined whether the respondents perceived the scent as being (in)congruent with the service experience (Bone and Jantrania 1992) on a seven-point bipolar scale (1 = fully disagree / 7 = fully agree), and their perception of its properties in terms of stimulation (arousal, four items, seven-point bipolar scales ranging from -3 to +3, Cronbach's $\alpha = .73$) and pleasantness (five items, seven-point bipolar scales ranging from -3 to +3, Cronbach's $\alpha = .95$) (Bosmans 2006; Fisher 1974). Table A1 presents all items and descriptive statistics.

Table A1. Items and results of the pretest in the field

Scent property	Item(s)	Scale type	Mean	Standard deviation	Cronbachs α
Congruence	The scent fits my ideal train ride experience	7-point Likert scale (1 = fully disagree/ 7 = fully agree)	4.04	1.85	n/a
Arousal	Relaxed/ tense Boring/ stimulating Unlively/ lively Dull/ bright	7-point bipolar scale (-3/+3)	0.22	0.96	.733
Pleasantness	Good/ bad Pleasurable/ unpleasurable Comfortable/ uncomfortable Positive/ negative Attractive/ unattractive	7-point bipolar scale (-3/+3)	0.75	1.42	.951
Familiarity	The scent is familiar to me.	7-point Likert scale (1 = fully disagree/ 7 = fully agree)	3.53	1.53	n/a

The consumers indicated that the scent is neutral congruent (mean 4.04) with an ideal train ride experience, because the mean item score did not significantly deviate from the scale mid-point of 4 ($t_{(47)} = .156$, $p = .876$, effect size $\eta^2 = .001$). The scent's arousal level was perceived as neutral and did not differ significantly from the scale midpoint of zero ($t_{(38)} = 1.418$, $p = .164$, $\eta^2 = .050$), further supporting the results obtained in the first lab pretest. Thus, the scent fits the consumers' target-arousal level (Wirtz, Mattila, and Tan 2000). Also in line with the first pretest, the scent was evaluated as pleasant, with a significant deviation from the scale midpoint of zero ($t_{(39)} = 3.348$, $p = .002$, $\eta^2 = .223$), which is a necessary condition for the scent to have positive effects in consumers' service evaluations.

When selecting a scent, we consider a scent's pleasantness to be the most important factor, even compared with its congruence, since individuals who receive repeated exposure will learn to associate the scent with their de facto train ride experience over time due to associative learning (Biswas et al. 2014; Degel, Piper, and Köster 2001; Epstein et al. 2009; Herz 2005).

Finally, we controlled for a low familiarity impression of the stimulus (seven-point bipolar scale, 1 = fully disagree/ 7 = fully agree) to avoid any impact of previous scent experiences (Morrin and Ratneshwar 2003). The respondents rated their familiarity as low, deviating significantly from its neutral midpoint ($t_{(44)} = -2.044$, $p = .047$, $\eta^2 = .086$). Overall, these results suggest that the scent stimulus is appropriate for our experiment.

Online Appendix B: Construct Measures across study waves in Study 2

Table B1. Constructs, and Quality Criteria across study waves (Study 2)

Construct	Item #	Factor Loading	KMO	Variance Explained	Cronbach's α	Discriminant Validity via HTMT criterion (Henseler, Ringle, and Sarstedt 2015)	Item-to-Total Correlation
		mean [min-max]	mean [min-max]	mean [min-max]	mean [min-max]	bootstrapped mean per study wave [bias-corrected 95% C.I.'s]	mean [min-max]
Perceived Service Quality (Dabholkar, Shepherd, and Thorpe 2000)	1	.912 [.859-.959]	.815 [.752-.871]	.861 [.811-.900]	.944 [.920-.961]	Service Experience / Service Value 1: .593 [.273-.808] / .491 [.234-.695] 2: .593 [.273-.808] / .734 [.452-.910] 3: .707 [.378-.919] / .691 [.428-.895] 4: .704 [.406-.916] / .716 [.446-.903] 5: .674 [.410-.848] / .752 [.516-.911] 6: .778 [.537-.920] / .842 [.620-.966] 7: .833 [.651-.939] / .792 [.449-.968] 8: .823 [.591-.944] / .785 [.539-.934] 9: .836 [.675-.960] / .902 [.682-.998]	.843 [.751-.923]
	2	.939 [.919-.964]					.888 [.847-.927]
	3	.943 [.915-.983]					.897 [.851-.969]
	4	.916 [.860-.937]					.851 [.758-.887]
Service Experience (Brady and Cronin 2001)	1	.881 [.773-.947]	.702 [.658-.763]	.823 [.670-.926]	.883 [.746-.957]	Service Quality: see row for Perceived Service Quality Service Value: see row for Perceived Service Value	.741 [.513-.882]
	2	.938 [.868-.976]					.846 [.662-.941]
	3	.898 [.812-.967]					.769 [.567-.923]
Perceived Service Value (Harris and Goode 2004)	1	.905 [.855-.946]	.807 [.758-.837]	.758 [.705-.797]	.891 [.856-.915]	Service Quality: see row for Perceived Service Quality Service Experience 1: .536 [.179-.658] 2: .572 [.288-.822] 3: .539 [.201-.785] 4: .627 [.292-.891] 5: .623 [.316-.834] 6: .581 [.293-.793] 7: .622 [.272-.869] 8: .587 [.246-.831] 9: .771 [.513-.953]	.816 [.731-.889]
	2	.744 [.583-.849]					.604 [.422-.723]
	3	.911 [.839-.952]					.826 [.709-.900]
	4	.907 [.836-.942]					.817 [.703-.877]

Notes: [min – max] describe the minimum and maximum values across all nine study waves.

Online Appendix C: Analysis of Panel Conditioning in Study 2

To control for panel conditioning, we introduced a control panel over three waves (1-3) and two further control groups (one-time measurement) on another track section in the same region (both without scent, see Figure 3 in the main article). By comparing the experimental and the control panel, we were able to rule out possible adverse arising from repeatedly interviewing the same consumers (i.e., mere measurement effects, Dholakia and Morwitz 2002).

The recruiting process, instructions, and questionnaire for the control panel were the same as for our main panel. The respondents received the questionnaire via mail and a ticket voucher worth EUR 30 after study completion. In total, 25 respondents (12 female, between 15 and 54 years old) participated in all three waves. In the absence of a panel conditioning effect, the control panel should not differ regarding the constructs under research between the waves. As expected, the construct service value did not change significantly over the three waves, which suggests that a practice effect is unlikely to occur in the main study (see Table C1, Panel A).

However, the respondents' assessment of service quality ($\chi^2_{(2)} = 7.586$, $p = .023$) and service experience ($\chi^2_{(2)} = 9.477$, $p = .009$) changed significantly over time. More precisely, we observed a positive deviation in service quality ($p = .059$, $\eta^2 = .109$), between waves 1 and 2, and a positive change in service experience between waves 1 and 3 ($p = .017$, $\eta^2 = .152$; Table C1, Panel B). These results indicate a potential panel conditioning. Consequently, we proceeded with further in-depth analysis, involving two additional control groups in the same trains in waves 2 ($n = 39$) and 3 ($n = 10$) (Figure 3 in the main article). This step allowed us to compare the mean values of the control panel answering the questionnaire for the second or the third time, with the control group respondents participating for the first time on the same track section and at the same points in time.

Table C1. Tests of Panel Conditioning (Study 2)

			Perceived Service Quality (Dabholkar, Shepherd, and Thorpe 2000)	**Service experience** (Brady and Cronin 2001)	**Perceived Service Value** (Harris and Goode 2004)
Panel A	N		25	25	25
	df		2	2	2
	χ^2_F		7.586	9.477	2.523
	p		**.023**	**.009**	.283
Panel B	Wave 1 vs. Wave 2	χ^2_F	-.660	-.600	
		Adj. p	**.059**	.102	
		η^2	.109	.090	
	Wave 1 vs. Wave 3	χ^2_F	-.600	-.780	
		Adj. p	.102	**.017**	
		η^2	.090	.152	
	Wave 2 vs. Wave 3	χ^2_F	.060	-.180	
		Adj. p	1.000	.525	
		η^2	.001	.008	
Panel C	Wave 2 control panel vs. control group (n =39)	U	416.500		
		p	.328		
		η^2	.015		
	Wave 3 control panel vs. control group (n = 10)	U		98.500	
		p		.339	
		η^2		.027	

Notes: Panel A shows the results of a series of Friedman's ANOVA by rank to show whether there are any significant differences between the waves (e.g., mere measurement or practicing effects). Panel B shows the results of post-hoc tests for all combinations of waves where the main effect was significant. In order to rule out potential panel conditioning effects, we then compared the control panel with the control group in the corresponding waves using Mann–Whitney U tests (Panel C).

Results from a Mann–Whitney U test indicate that the service quality levels did not differ significantly between the two groups (wave 2: $U = 416.50$, $p = .328$, $\eta^2 = .015$). In addition, the service experience evaluations did not deviate significantly between the two groups (wave 3: $U = 98.50$, $p = .339$, $\eta^2 = .027$; Table C1, Panel C). We conclude that our results are not biased by panel conditioning effects due to mere measurement effects.

REFERENCES

Biswas, Dipayan, Courtney Szocs, Aradhna Krishna, and Donald R. Lehmann (2014), "Something to Chew On: The Effects of Oral Haptics on Mastication, Orosensory Perception, and Calorie Estimation," *Journal of Consumer Research*, 41 (2), 261–73.

Bone, Paula F. and Swati Jantrania (1992), "Olfaction as a Cue for Product Quality," *Marketing Letters*, 3 (3), 289–96.

Bosmans, Anick (2006), "Scents and Sensibility: When Do (In)Congruent Ambient Scents Influence Product Evaluations?," *Journal of Marketing*, 70 (3), 32–43.

Brady, Michael K. and Joseph J. Cronin (2001), "Some New Thoughts on Conceptualizing Perceived Service Quality: A Hierarchical Approach," *Journal of Marketing*, 65 (3), 34–49.

Dabholkar, Pratibha A., C. David Shepherd, and Dayle I. Thorpe (2000), "A Comprehensive Framework for Service Quality: An Investigation of Critical Conceptual and Measurement Issues Through a Longitudinal Study," *Journal of Retailing*, 76 (2), 139–73.

Degel, Joachim, Dag Piper, and Egon P. Köster (2001), "Implicit Learning and Implicit Memory for Odors: The Influence of Odor Identification and Retention Time," *Chemical Senses*, 26 (3), 267–80.

Dholakia, Utpal M. and Vicki G. Morwitz (2002), "The Scope and Persistence of Mere-Measurement Effects: Evidence from a Field Study of Customer Satisfaction Measurement," *Journal of Consumer Research*, 29 (2), 159–67.

Doucé, Lieve, Karolien Poels, Wim Janssens, and Charlotte de Backer (2013), "Smelling the Books: The Effect of Chocolate Scent on Purchase-related Behavior in a Bookstore," *Journal of Environmental Psychology*, 36 (December), 65–9.

Epstein, Leonard H., Jennifer L. Temple, James N. Roemmich, and Mark E. Bouton (2009), "Habituation as a Determinant of Human Food Intake," *Psychological Review*, 116 (2), 384–407.

Fisher, Jeffrey D. (1974), "Situation-specific Variables as Determinants of Perceived Environmental aesthetic Quality and Perceived Crowdedness," *Journal of Research in Personality*, 8 (2), 177–88.

Harris, Lloyd C. and Mark M.H. Goode (2004), "The Four Levels of Loyalty and the Pivotal Role of Trust: A Study of Online Service Dynamics," *Journal of Retailing*, 80 (2), 139–58.

Henseler, Jörg, Christian M. Ringle, and Marko Sarstedt (2015), "A New Criterion for Assessing Discriminant Validity in Variance-based Structural Equation Modeling," *Journal of the Academy of Marketing Science*, 43 (1), 115–35.

Herz, Rachel S. (2005), "Odor-associative Learning and Emotion: Effects on Perception and Behavior," *Chemical Senses*, 30 Suppl 1, i250-1.

Krishna, Aradhna, May O. Lwin, and Maureen Morrin (2010), "Product Scent and Memory," *Journal of Consumer Research*, 37 (1), 57–67.

Morrin, Maureen and S. Ratneshwar (2003), "Does It Make Sense to Use Scents to Enhance Brand Memory?," *Journal of Marketing Research*, 40 (1), 10–25.

Spangenberg, Eric R., Ayn E. Crowley, and Pamela W. Henderson (1996), "Improving the Store Environment: Do Olfactory Cues Affect Evaluations and Behaviors?," *Journal of Marketing*, 60 (2), 67–80.

Wirtz, Jochen, Anna S. Mattila, and Rachel L.P. Tan (2000), "The Moderating Role of Target-arousal on the Impact of Affect on Satisfaction: An Examination in the Context of Service Experiences," *Journal of Retailing*, 76 (3), 347–65.