WEB APPENDICES

Crossing incentive alignment and adaptive designs in choice-based conjoint: a fruitful endeavor

Verena Sablotny-Wackershauser, Marcel Lichters, Daniel Guhl, Paul Bengart, Bodo Vogt

Table of contents

Appendix A: Additional theoretical background

This section provides additional information about (1) mechanisms for incentive alignment (Tables A1 and A2), (2) the popularity of adaptive designs in the literature (Table A3), (3) the options to combine different stages in ACBC studies (Table A4), and (4) specific considerations for combining incentive alignment and adaptive designs (Table A4).

Mechanisms for incentive alignment and their boundaries

Table A1. Mechanisms to incentive-align CBC

	Direct Mechanism Ding et al. (2005)	**Inferred WTP** Ding (2007)	**Inferred RankOrder** Dong et al. (2010)	**Pay-for-Performance** Sipos and Voeth (2015)	**Award Mechanism** Vadali (2016)
Action by partici-pant	Participant completes CBC tasks	Participant completes conventional conjoint (e.g. CBC)	Participant completes conventional conjoint (e.g. CBC)	Participant completes tasks in CBC and HOTs	Participant completes CBC tasks
Action by analyst	1. Analyst evaluates ANSWERS TO CHOICE SETS	1. Analyst reveals ONE REAL PRODUCT X	1. Analyst reveals TWO OR MORE REAL PRODUCTS X_i	1. Analyst determines the NUMBER OF FIRST CHOICE HITS in HOTs	1. Analyst DIVIDES CBC CHOICE TASKS into estimation and validation sets
	2. Analyst selects one of participant's choice tasks randomly	2. Analyst infers participant's WTP for real product X on the basis of the conjoint task	2. Analyst infers participant's relative preferences (rank order) for real products X_i on the basis of the conjoint task	2. Analyst calculates the monetary reward dependent on the number of first choice hits	2. Analyst estimates partworth utilities based on estimation set and calculates first choice hits for the validation set
Outcome	Participant receives preferred option from the randomly chosen choice task	Actual purchase price (x) determined by an incentive-aligned mechanism (BDM). Then, compared with inferred WTP: Participant might purchase product	Participant receives the top-ranked product based on inferred preference order	Participant receives his or her monetary reward (fixed + variable or just variable reward)	Participant for whom the prediction is best receives monetary disbursement (competition context)
Variants	**Progressive Direct Mechanism** Eggers (2012)		**Progressive RankOrder** Eggers (2012)		**Hybrid Award + Inferred WTP** Vadali (2016)

Table A2. Application boundaries of incentive alignment mechanisms

	Incentive alignment using				
	Direct mechanism	**Inferred WTP***	**Inferred RankOrder**	**Pay-for-Performance**	**Award mechanism***
Reward specification	direct	indirect	indirect	indirect	indirect
Immediate incentivization	easy	cumbersome (utility estimation first)	cumbersome (utility estimation first)	cumbersome (utility estimation first)	cumbersome (utility estimation first)
Incentivization is affected by estimate accuracy	no	yes	yes	yes	yes
Product as reward	yes	yes	yes	no	no
Price as a necessary attribute	no	yes	no	no	no
Availability of products	all	one	at least two	none	none
Comprehensib ility	+	-	+	0	0
Additional tasks for participants	none	additional BDM task	none	additional HOTs	none
Mandatory no-purchase option	yes, in studies without monetary endowment when the price is involved	yes	yes, in studies without monetary endowment when the price is involved	*free-choice design not advisable: mechanism may be undermined by consistently selecting the no-purchase option*	

* Vadali's (2016) hybrid version combines both procedures (Award + inferred WTP)

Approaches for adaptive CBC designs

Table A3. Popularity of adaptive CBC designs

Adaptive CBC Design	Applications with Methodological Focus	Applications with Content Focus[a]	Cum. (Mean) Impact Factor[b]	Original Source
Adaptive Choice-based Conjoint (ACBC)	Bacon and Lenk (2012) Bauer et al. (2015) Boesch et al. (2013) Böhm et al. (2013) Brand and Baier (2020) Buoye et al. (2018) Chapman et al. (2009) Cunningham et al. (2010) Gundlach and Hofmann (2020) Huang and Luo (2016) Isihara et al. (2020) Orme (2009b) Orme and Johnson (2008)	Almario et al. (2021) Al-Omari (2017) Al-Omari and McMeekin (2020) Al-Omari et al. (2017) Boesch (2012) Boyer et al. (2021a) Boyer et al. (2021b) Brand and Rausch (2021) Brand et al. (2022) Brodock et al. (2021) Burs et al. (2020) Cahalan and Faber (2015) Chassot et al. (2014) Chowdhury et al. (2016) Cocquyt et al. (2020) Cook et al. (2019) Côté and Salm (2022) Cunningham et al. (2015) De Groot et al. (2011; 2012) Deal (2014) Del Toro-Gipson et al. (2021) Deshwal et al. (2022) Feilhauer et al. (2022) Flöthmann et al. (2018) Fuchs and Hovemann (2022) Giessmann et al. (2015) Gundlach and Hofmann (2020) Gutierrez et al. (2021) Gutierrez et al. (2022) Habann et al. (2018) Haggenmüller et al. (2021) Harwood and Drake (2018; 2019; 2020) Hinnen et al. (2017) Holopainen et al. (2020) Hoppe et al. (2022) Hu et al. (2022) Hunka et al. (2021) Inci et al. (2022) Issa et al. (2013) Jazdzik-Osmolska (2021) Jervis et al. (2012; 2014b) Jervis et al. (2014a) Joo et al. (2020) Khan et al. (2020) Kim et al. (2013) Klein et al. (2019) Klein et al. (2021) Knoerl et al. (2022) Kouki-Block and Wellbrock (2021) Krewson and Owens (2022) Lima et al. (2020) Lima et al. (2022) Lysaght et al. (2021) Mai et al. (2022) McCarty et al. (2017) McLean et al. (2017) Melliger and Lilliestam (2021) Mengelkamp et al. (2019)	511.308 (5.382)	Johnson and Orme (2007)

Adaptive CBC Design	Applications with Methodological Focus	Applications with Content Focus[a]	Cum. (Mean) Impact Factor[b]	Original Source
		Naous et al. (2022) Nickkar and Lee (2022) Oltman et al. (2014; 2015) Petrovich et al. (2019) Pierce et al. (2020) Pleger et al. (2020) Racette and Drake (2022) Reinisch et al. (2021) Rivera et al. (2021) Rizzo et al. (2020) Ronda et al. (2021) Salm (2017) Salm et al. (2016) Salm and Wüstenhagen (2018) Schomakers et al. (2021) Schomakers and Ziefle (2022) Schwoerer et al. (2020) Shepperson et al. (2016) Shin et al. (2015) Siegel et al. (2016) Silver (2018) Sipple et al. (2022) Smith et al. (2016) Speed et al. (2022) Speight et al. (2019) Stone et al. (2021) Stone et al. (2022) Storm-Dickerson et al. (2018) Sulistyawati et al. (2020) van den Borne et al. (2017) Veitch et al. (2017) Veitch et al. (2021) Veitch et al. (2022) Wicki and Kaufmann (2022) Will et al. (2022) Wuebker et al. (2015) *Expected increase in practical relevance; 18% of projects on WTP estimation (Steiner and Hendus 2012); 13% of projects considering Sawtooth users (Sawtooth Software 2019b)*		
Polyhedral Adaptive CBC	Abernethy et al. (2005, 2008) Bertsimas and O'Hair (2013) Eggers (2008) Gaskin et al. (2007) Joo et al. (2019) Sauré and Vielma (2019)	*Expected a decrease in practical relevance (Steiner and Hendus 2012)*	38.8 (5.114)	Toubia et al. (2003; 2004; 2007)
Separated Adaptive Dual Response (SADR)		Heidel et al. (2021) Lehmann et al. (2022) Rheindorf et al. (2021)	30.0 (7.5)	Schlereth and Skiera (2017)
Individually Adapted Sequential (Bayesian) Designs	Consonni et al. (2020) Crabbe et al. (2014) Danthurebandara et al. (2011, 2015) Traets et al. (2020)		20.1 (4.02)	Yu et al. (2009, 2011)
Hessian-based Adaptive CBC	Huang and Luo (2016) Ren and Scott (2017) Toubia et al. (2013)		19.3 (6.433)	Abernethy et al. (2005, 2008)

Adaptive CBC Design	Applications with Methodological Focus	Applications with Content Focus[a]	Cum. (Mean) Impact Factor[b]	Original Source
Individually Adapted CBC		Ebbers et al. (2021) Lemken (2021)	18.6 (6.2)	Gensler et al. (2012)
Hybrid Individualized Two-level CBC	Bauer et al. (2015)	Ronda et al. (2021) Zenker et al. (2013)	18.3 (6.1)	Eggers and Sattler (2009)
Active Machine Learning	Huang and Luo (2016)		10.0 (5.0)	Dzyabura and Hauser (2011)
Compound Design Criterion for Selective Choice Proc.			6.1 (6.1)	Henderson and Liu (2017)
Goal-directed Question-Selection Method			5.4 (5.4)	Joo et al. (2019)
Fuzzy SVM Method			5.411 (5.411)	Huang and Luo (2016)/ Kang et al. (2019)
Bayes-Optimal Entropy Pursuit for Choice-based Preference Elicitation			5.0 (5.0)	Pallone et al. (2017)
Ellipsoidal Method for Adaptive CBC			2.7 (2.7)	Sauré and Vielma (2019)
Fuzzy Learning and Clustering			2.5 (2.5)	Banerjee et al. (2011)

Notes: The literature review includes articles published until the end of 2022. It omits working papers, conference papers, and Ph.D. and master theses. The adaptive CBC designs are sorted in decreasing order of their cumulative impact factor. Columns two and three exclude the original sources.
[a] publications in the fields of health research, marketing and advertising, food studies, transportation research, and energy and environmental research, as well as entrepreneurship studies (ACBC search canceled on the 1000th Google Scholar hit)
[b] based on journal impact factors according to the 2022 Journal Citation Reports of Clarivate Web of Science. Publications without an impact factor were excluded from the calculation of the mean impact factor.

Based on the systematic literature review, we identify and briefly describe 14 different adaptive CBC approaches (or their respective methods) – listed in Table A3 in the decreasing order of their cumulative impact factor. Since *ACBC* following Johnson and Orme (2007) is described in detail in the main article to which this Web Appendix refers, we forego describing its procedure and start with the second adaptive CBC design.

The *Polyhedral Adaptive CBC*, introduced by Toubia et al. (2004), is an adaptive question-selection procedure that quickly reduces the sets' relevant attribute levels to individual sets consistent with the participant's choices, portraying possible partworth vectors as a polyhedron. To date, however, this approach has no freely available software application (Baier and Brusch 2021). This might hinder its widespread adoption. Furthermore, as Steiner and Hendus (2012) have found, practitioners from 34 market research companies expect *Polyhedral Adaptive CBC* to be less relevant in the future.

The *Separated Adaptive Dual Response (SADR)* approach, developed by Schlereth and Skiera (2017), seeks to improve the elicitation of each participant's WTP and buy/no-buy-indifference threshold. It is a modification of the dual response choice design (Brazell et al. 2006) and comprises two separate survey blocks: the forced-choice block and the free-choice block. This approach uses the responses given in the (static) first block to adaptively optimize the second block by selecting fewer but more informative questions.

The *Individually Adapted Sequential (Bayesian)* approach proposed by Yu et al. (2011) continuously updates prior information for each participant after each response, generates individualized choice sets that maximize D-optimality at the individual level. Thus, this approach tries to increase the precision of estimated main effects. According to its authors, it is more robust to response errors than *ACBC* and *Polyhedral Adaptive CBC*.

The *Hessian-based Adaptive CBC* is based on minimizing the loss function, instead of maximizing some measure of fit as is common in traditional estimation approaches such as *Polyhedral Adaptive CBC* (Abernethy et al. 2008). The advantage of this approach is that it is robust to response errors (i.e., the participant did not always choose the utility-maximizing option). Like the approaches described above, we found only method-driven literature on *Hessian-based Adaptive CBC*. The approach is portrayed as a generalization of *Polyhedral Adaptive CBC*, and it also comes without a readily available software package. However, as an advantage over *Polyhedral Adaptive CBC*, this approach never ends up in a situation where the analytic center of the polyhedron does not converge on the true partworths of a participant.

Unfortunately, to date, applications of this approach to free-choice CBC, including a no-choice option, are absent, which limits its use for CBC accompanied by, among others, market simulations.

The *Individually Adapted CBC* focuses on the adaption of price levels (Gensler et al. 2012). Based on the participant's behavior, an algorithm continuously adapts the prices either upward or downward until the price reaches the participant's willingness-to-pay (WTP). This easy-to-implement approach aims to prevent extreme response behavior, such as always or never choosing the no-buy option, which can occur when there is no overlap between the price levels and the participants' WTP. Paradoxically, this strength is also a weakness of this approach, since it is only applicable if the price plays a role in the conjoint analysis design, which is not always the case.

The *Hybrid Individualized Two-level CBC* proposed by Eggers and Sattler (2009) implements choice designs consisting only of the best and worst levels of attributes, based on the results of a preceding, self-explicated preference measurement approach. Because this approach involves only two levels per attribute, it has the advantage of being both easier to design and insensitive to possible bias resulting from the number-of-levels effect (Currim et al. 1981). Furthermore, the approach allows, to a certain extent, for an a priori exclusion of choice tasks that would present dominated alternatives. A noted drawback of this approach is that attribute level utilities between the best and worst levels of any attribute are derived from rating scales instead of comparative decisions. Furthermore, researchers must specify possible interactions between different attributes prior to fielding a study.

Dzyabura and Hauser's (2011) *Active Machine Learning* approach is based on the authors' proposed Bayesian algorithm to update, iteratively, prior and posterior beliefs about the participants' heuristic decision rules. This approach distinguishes between the non-compensatory screening of alternatives and preference ordering. In addition, it allows the use of data obtained from previous participants to inform the priors for subsequent data collection.

The next adaptive CBC approach is the *Compound Design Criterion for Selective Choice Process* (Henderson and Liu 2017). It is a Bayesian approach that integrates participants' self-stated information on whether specific attributes matter to them (or not). The information is collected in the initial stage of the study and is then used to customize individual-specific choice tasks: the attributes that matter receive more weight, and those that do not matter receive less weight. However, the approach reaches its limits when two or more attributes only gain specific importance in combination (i.e., interaction effects).

The *Goal-directed Question-Selection Method* (Joo et al. 2020) is one of the most recent approaches listed in Table A3 and was introduced to manage conjoint studies with high dimensionality (i.e., many attributes and/or levels per attribute). It uses a question-selection criterion that selects product profiles with high uncertainty in high-preference regions to achieve a balance between high-performance products and high-potential products.

The overall goal of the method is not to select choice tasks that adaptively optimize the precision of parameter estimates, but to select choice tasks that will uncover attribute combinations (including interaction effects) with a high potential to achieve the highest possible market share. Thus, the method is inspired by product development decisions that are usually considered by management. It allows for the integration of a no-purchase option.

The *Fuzzy SVM Method* was developed to handle preference measurement involving complex product concepts with many attributes and/or attribute levels (Huang and Luo 2016; Kang et al. 2019). Like ACBC, the proposed method begins with a product configurator task and allows participants to indicate must-have and unacceptable features. In addition, like the *Active Machine Learning* approach (Dzyabura and Hauser 2011), it permits the use of the responses of previous participants.

As reflected in its name, Pallone et al.'s (2017) *Bayes-Optimal Entropy Pursuit for Choice-based Preference Elicitation* approach uses a greedy algorithm to minimize the posterior entropy of linear classifiers that predict participants' preferences, which provides the basis for the adaptive selection of questions. Interestingly, the authors explicitly formulate CBC as a finite horizon, adaptive learning problem. However, applying the proposed approach to a real-world CBC problem or even a simulation study on consumers' product choices is nonexistent.

Sauré and Vielma (2019) developed the *Ellipsoidal Method for Adaptive CBC* to overcome the shortcoming of Toubia et al.'s (2004) *Polyhedral Adaptive CBC*, namely poorer performance when response error is high, while preserving the simplicity. This method updates the ellipsoidal credibility region after each question while considering the possibility of response errors.

Fuzzy Learning and Clustering (Banerjee et al. 2011) is the last adaptive CBC approach listed in Table A3. As its name indicates, this approach uses a fuzzy logic-based algorithm to generate questions based on the participant's previous responses. According to its authors, the non-discrete nature of the applied fuzzy logic makes the question generation process more reliable and efficient.

Although many different approaches are available to generate and design adaptive CBCs, only three of them – *ACBC*, *Hybrid Individualized Two-level CBC*, and *Individually Adapted CBC* – have been applied in content-focused research. By comparing the number of references listed for each approach, it is apparent that *ACBC* is by far the most widely used adaptive CBC approach. When taking a closer look at these references, it is obvious that *ACBC* has been applied in various management fields to elicit preferences for, among others, food (e.g., Brodock et al. 2021; McCarthy et al. 2017), health-related services and products (e.g., Cook et al. 2020; Reinisch et al. 2021), energy (projects) (e.g., Melliger and Lilliestam 2021; Mengelkamp et al. 2019), infrastructure services (e.g., Lima et al. 2020; Silver 2018), job attributes (Ronda et al. 2021), and managerial competences (Flöthmann et al. 2018). The widespread use and popularity of *ACBC* in various disciplines are the main reasons why we chose *ACBC* over other adaptive CBC approaches to investigate the effect of incentive alignment, as we aim to disclose implications relevant to a broad spectrum of researchers and practitioners.

Considerations for implementing shortened or incentive-aligned ACBC designs

Table A4 aims to assist researchers and analysts in determining whether to exclude ACBC stages from the preference elicitation interview flow and how this decision impacts the process and outcomes of an ACBC study.[1] The table furthermore illustrates the available mechanisms for potential incentive alignment for each of the shortened ACBC versions (see the last three columns). For this purpose, we solely focus on the three most prominent mechanisms in the marketing research literature (see Table 1 in the manuscript and also Tables A1 and A2 for more details). These include the Direct Mechanism (Ding et al. 2005), the RankOrder mechanism (Dong et al. 2010), and the inferred WTP mechanism (Ding 2007).

[1] We thank the anonymous review team for these suggestions.

Table A4. Assessment of permutations of ACBC stages and possible incentive aligning mechanisms

	ACBC Stage						Incentive alignment mechanism		
	I	**II**	**III**	**IV**	**Feasible/ advisable**	**Consequences for preference elicitation**	**Direct mechanism**	**RankOrder**	**Inferred WTP**
A	✓	✓	✓	✓	doable	Longest interview time of all variants. However, on average, superior predictive validity according to our empirical investigations.	Yes (winner *Choice Tournament*)	Yes	Yes
B	✓	✓	✓	✗	doable	No fine-tuning of the no-purchase option's/outside good's utility.	Yes (winner *Choice Tournament*)	Yes	Yes
C	✓	✓	✗	✗	doable	No observations for direct trade-offs among product concepts that belong to a consumer's consideration set.	No	Yes	Yes
D	✓	✓	✗	✓	doable	*Calibration* stage constrained, because it usually needs information from the *Choice Tournament* (e.g., the tournament winning concept). No observations for direct trade-offs among product concepts that belong to a consumer's consideration set.	No	Yes	Yes
E	✓	✗	✗	✗	Doable, but: no information on the outside good's utility. No information on must-have and must-avoid attribute levels.	Very sparse information at the individual level, which comes with smoothing toward population means in HB estimation (obfuscating consumer heterogeneity). No predictions of markets that provide the option not to buy at all. No information on non-compensatory decision strategies (e.g., for further consumer segmentation). No observations for direct trade-offs among product concepts that belong to a consumer's consideration set.	No	Yes	No
F	✓	✗	✓	✗	Not advisable, because concepts that do not belong to an individual's consideration set might be included in the (forced-)*Choice Tournament* stage (i.e., consumers must choose among irrelevant alternatives). No information on the outside good's utility. No information on must-have and must-avoid attribute levels.	No predictions of markets that provide the option not to buy at all. No information on non-compensatory decision strategies (e.g., for further consumer segmentation). Consumers may face ample irrelevant concepts in the *Choice Tournament*.	No, because one cannot be sure that *Choice Tournament* winner is truly in consideration set.	Yes	No
G	✓	✗	✓	✓	Not advisable, because concepts that do not belong to an individual's consideration set might be included	*Calibration* stage constrained, because it usually needs information from screening (i.e., concepts that are clearly above or below an individual's utility for the	No, because one cannot be sure that *Choice*	Yes	Yes

	ACBC Stage						Incentive alignment mechanism		
	I	**II**	**III**	**IV**	**Feasible/ advisable**	**Consequences for preference elicitation**	**Direct mechanism**	**RankOrder**	**Inferred WTP**
					in the (forced-)*Choice Tournament* stage (i.e., consumers must choose among irrelevant alternatives). No information on must-have and must-avoid attribute levels.	outside good). Consumers may face ample irrelevant concepts in the *Choice Tournament*. No information on non-compensatory decision strategies (e.g., for further consumer segmentation).	*Tournament* winner is truly in consideration set.		
H	✓	✗	✗	✓	Doable, but: no information on must-have and must-avoid attribute levels.	Very sparse information at the individual level, which comes with smoothing toward population means in HB estimation (obfuscating consumer heterogeneity). No information on non-compensatory decision strategies (e.g., for further consumer segmentation). No observations for direct trade-offs among product concepts that belong to a consumer's consideration set.	No	Yes	Yes
I	✗	✓	✓	✗	doable	No direct observation of an individual's "dream concept" (might give an impression on how consumers might behave when using product configurators). *Screening* stage is expected to last longer, since no good starting point for near-optimal concepts is provided.	Yes (winner *Choice Tournament*)	Yes	Yes
J	✗	✓	✓	✓	doable	No direct observation of an individual's "dream concept" (might give an impression on how consumers might behave when using product configurators). *Screening* stage is expected to last longer, since no good starting point for near-optimal concepts is provided.	Yes (winner *Choice Tournament*)	Yes	Yes
K	✗	✓	✗	✗	doable	No direct observation of an individual's "dream concept" (might give an impression on how consumers might behave when using product configurators). *Screening* stage is expected to last longer, since no good starting point for near-optimal concepts is provided. No observations for direct trade-offs among product concepts that belong to a consumer's consideration set.	No	Yes	Yes
L	✗	✓	✗	✓	doable	No direct observation of an individual's "dream concept" (might give an impression on how consumers might behave when using product configurators).	No	Yes	Yes

	ACBC Stage						Incentive alignment mechanism		
	I	II	III	IV	**Feasible/ advisable**	**Consequences for preference elicitation**	**Direct mechanism**	**RankOrder**	**Inferred WTP**
						Screening stage is expected to last longer, since no good starting point for near-optimal concepts is provided. No observations for direct trade-offs among product concepts that belong to a consumer's consideration set. *Calibration* stage constrained, because it usually needs information from the *Choice tournament* (e.g., the tournament winning concept).			
M	✗	✗	✓	✗	Not advisable, because concepts that do not belong to an individual's consideration set might be included in the (forced-)*Choice Tournament* stage (i.e., consumers must choose among irrelevant alternatives). No information on must-have and must-avoid attribute levels.	Would approximate a classical CBC in a forced-choice design. No direct observation of an individual's "dream concept" (might give an impression on how consumers might behave when using product configurators). No information on non-compensatory decision strategies (e.g., for further consumer segmentation).	No, because one cannot be sure that *Choice Tournament* winner is truly in consideration set	Yes	No
N	✗	✗	✗	✓	not feasible, because *Calibration* needs information on product concepts' utilities as input.		No	No	No
O	✗	✗	✓	✓	Not advisable, because concepts that do not belong to an individual's consideration set might be included in the (forced-)*Choice Tournament* stage (i.e., consumers must choose among irrelevant alternatives). No information on must-have and must-avoid attribute levels.	No direct observation of an individual's "dream concept" (might give an impression on how consumers might behave when using product configurators). No information on non-compensatory decision strategies (e.g., for further consumer segmentation). A proper function of the *Calibration* stage is not quarantined, because even using "winners" and "losers" from the *Choice Tournament* constitute the input for the *Calibration* stage, it remains unclear whether they provide utility below/ or above the outside good.	No, because one cannot be sure that *Choice Tournament* winner is truly in consideration set	Yes	Yes

Notes: I – Build-your-own (*BYO*) stage, II – *Screening* stage, III – *Choice Tournament* stage, IV – *Calibration* stage, ✓, ✗ – implemented/ not implemented.

The overview provides several key insights. First, when considering the implementation of shortened ACBC versions – probably to reduce the overall interview time – combinations that simultaneously exclude the *Screening* and the *Calibration* stage are not suitable for many market research questions. These variants fail to provide an estimate for the utility of the outside option (i.e., the none parameter), which is essential for managerial relevant results such as simulations of general product demand, WTP, and valid estimates of (cross-)price elasticities. Second, certain combinations of stages are not feasible (e.g., N), while others (e.g., F and G) are not advisable because they may undermine the fundamental principles of ACBC. For example, removing the *Screening* stage forces consumers to choose among alternatives in the *Choice Tournament* that may not be relevant at all, compromising the essence of ACBC. Third, other combinations are feasible but will lead to the omission of interesting information that analysts commonly seek to extract from an ACBC study (i.e., non-compensatory decision rules will not be uncovered when dropping the *Screening* stage, a consumer's idiosyncratic preference in a product configurator will not be recorded when dropping the *BYO* stage). Fourth, when considering excluding ACBC stages, we advise starting to exclude them from the end. This is because all ACBC stages are interconnected. Thus, interesting aspects of ACBC in later stages (e.g., ensuring an informative set of product concepts to be included in the *Calibration* stage, making sure that each product concept is part of the participant's consideration set) will be less clear when excluding preceding stages. Finally, excluding single stages in ACBC studies makes it more challenging to uncover unobserved heterogeneity. That is, in a hierarchical Bayes framework, more shrinkage towards the population means can be expected. The reduced amount of data collected in shortened ACBC interviews limits the ability to capture individual-level variations.

In conclusion, it seems like there is no rose without a thorn when dropping ACBC stages. Consequently, market researchers should carefully evaluate if saved interview times truly compensate for the above drawbacks.

If the analyst decides to implement a shortened version of ACBC, it also has implications for incentivizing the preference elicitation task (last three columns in Table A3). Therefore, we now discuss the options available for incentivizing individual ACBC stages. First, dropping ACBC stages while selectively incentivizing certain stages, such as the *BYO* stage but not the *Screening* and *Choice Tournament* stages, can create a strong mental contrast for participants. By explicitly informing them that only the incentivized *BYO* stage is relevant for their payment, but not the other stages, it may affect their perception and engagement with the rest of the interview (this applies to other proposed variants as well). Typically, what is not measured is not managed, so we anticipate that exclusively incentivizing the *BYO* stage (or other stages) may not yield incremental positive effects on predictive validity, or it may even have adverse consequences. Participants are likely to allocate their effort primarily to the parts of the interview that offer incentives, potentially leading them to rush through the remaining parts (Yang et al. 2018).

Second, even when considering incentive alignment for single ACBC stages, further issues will remain. The strongest incentive scheme is given by the application of the *Direct mechanism* in ACBC. That is why we have chosen this mechanism for three of our four studies (i.e., Study 1: Pizzas, Supplemental Study A: PlayStation 4 Bundles, and Study 2: Food Processors). However, when only incentivizing, for example, the *Choice Tournament*, participants might not pay attention to the *BYO* and the *Screening* stage. As a result, the *Choice Tournament* might include irrelevant alternatives that are not part of the consideration set. In the *Choice Tournament,* participants cannot indicate that none of the presented products are worth buying for them, which would create substantial reactance when the *Direct mechanism* randomly selects one of the preferred products in the *Tournament* or the *Choice Tournament* winner as the study disbursement. The same line of reasoning applies when only incentivizing the *BYO* stage. Participants here are asked to configure their ideal product given the presented attribute levels and feature prices. They do not have the opportunity to indicate that the final configuration is truly worth buying for them, rendering the *BYO* configuration to be a suboptimal study disbursement. Similarly, it is not advisable to incentivize only the *Screening* stage or the combination of the *BYO* and *Screening* stages. In the *Screening*, stage participants indicate in each task which of multiple product alternatives are generally considered buying options and which are not (i.e., whether they would consider a product). However, no information is provided regarding participants' preferences among the alternatives within each task or whether they would actually purchase a product that was marked as a possible purchase option. Consequently, the *Direct mechanism* is only meaningfully applicable in ACBCs that provide at least the *Screening* and the *Choice Tournament* stage.

Third, as the *RankOrder mechanism* does not require a direct "purchase" decision from the participants, this class of incentive alignment is more versatile for situations in which an ACBC study does not include all stages. All that is needed is a preferably precise estimate of the total utility that the product concepts provide, which are included in the reward list of the mechanism. At the same time, one should keep in mind that this is the limitation of applying the *RankOrder* mechanism to shortened ACBC versions. After all, a precise estimate of the products' total utility is needed at the individual-level. If the participant, however, provides less information (as compared to a complete ACBC interview flow), then the likelihood increases that a suboptimal product is drawn as study disbursement.

Lastly, the inferred WTP mechanism relies on the availability of precise estimates for both the reward product's total utility and the utility of the outside good. This is because the study disbursement is determined through the BDM mechanism (Becker et al. 1964). A participant's

WTP is simulated based on comparing product utility with the utility of the outside good (Miller et al. 2011). Consequently, this incentive scheme is applicable only when a shortened ACBC includes the *Screening* and/or *Calibration* stage.

Appendix B: Supplemental Study A: Consumers' preferences for PlayStation 4 bundles

Preliminary study

As there is a vast amount of PS4 games available on the market, it was first necessary to select a useful range of current and popular games belonging to well-known game genres to be part of the pre-study. For this reason, we consulted several websites that provide online rankings of already launched games (e.g., metacritic.com) and those providing online announcements and test reports of games, which were supposed to be released during the period of research. We then cross-checked the shortlisted games with ratings on amazon.com and removed available games from the list, which did not reach at least three of five stars in the Amazon evaluation. After conducting additional ten informal inquiries with students and employees at a German university, we designed a preliminary online study using anchored MaxDiff scaling, also known as case I of Best-Worst Scaling (Louviere and Woodworth 1991b; Louviere et al. 2013) to select PS4 game genres, games, and accessories most relevant to the target group (similar to Study 1). Anchored MaxDiff, in contrast to unanchored MaxDiff, allows not only to estimate interval-scaled preference scores for the games under research but also to estimate the outside good's utility, enabling additional extraction of purchase probabilities (Lattery 2010; Orme 2009a). All participants completed 22 *Best-Worst tasks* with four simultaneously presented PS4 games (see Figure B1). Additionally, we asked in an open question-format whether participants think that an important game is missing in the MaxDiff study.

In total, 142 participants were recruited via online convenient sampling to take part in the survey (33% females, $M_{\text{age}} = 25.62$ years [$SD_{\text{age}} = 7.22$], 61% students). We raffled an €20 Amazon voucher among all participants to encourage participation. All of them passed through a screening procedure ensuring that they (1) had sufficient interest in the product category under research, (2) were willing to spend about 10 minutes on an online survey, (3) were at an age of at least 16 years, and (4) oversaw the purchase decision in their household for a PlayStation 4 inclusive PS4 games.

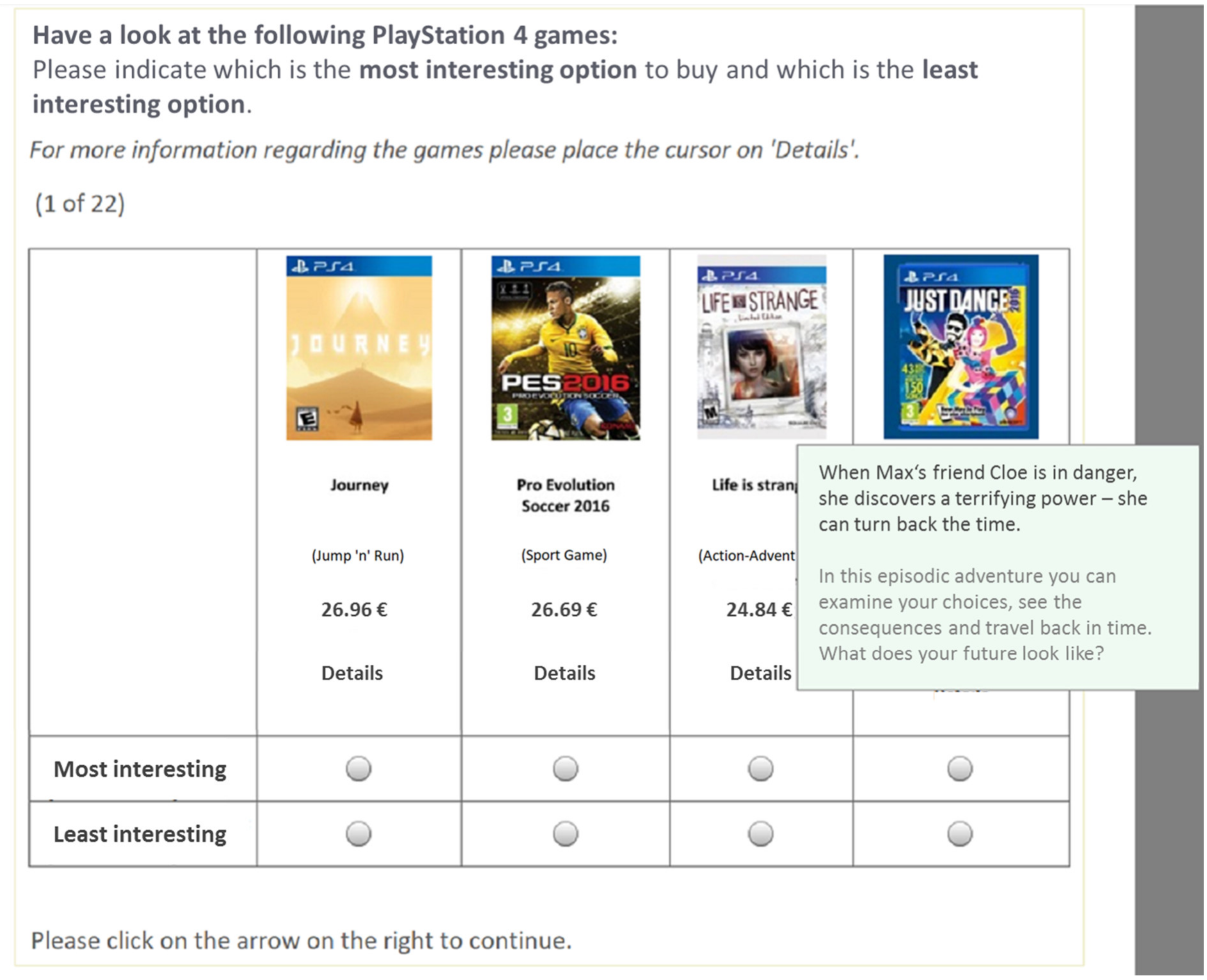


Fig. B1. Exemplary MaxDiff task in the preliminary study of the PS4 study (translated)

In the analysis, we focused particularly on the comparison of genres, applying Hierarchical Bayes (HB), Latent Class and Total Unduplicated Reach and Frequency (TURF) analyses (Miaoulis et al. 1990; Lagerkvist et al. 2012) to select three game genres to be included in the main study. Results of the Latent Class Analysis, especially the consistent Akaike Information Criterion (cAIC), also suggest a three-cluster-solution, concluding that each of the three selected game genres should address one cluster. Table A2 provides an overview of all pre-study results.

Table B1. Overview over the main findings in the MaxDiff survey on PS4 games

n = 142	MaxDiff counts			HB resc. prob. scores			Latent class			TURF analysis results							
	Count proportion				CI 95%		Solution for 3 segments			Ranking according		Share of preference	CI 95%		First Choice	CI 95%	
	Best	Worst	Diff.	Average	Lower	Upper	I 35.4%	II 28.0%	III 36.6%	Share of preference	First Choice	Average	Lower	Upper	Average	Lower	Upper
ACTION-ADVENTURE	**0.39**	**0.11**	**0.29**														
Far Cry Primal	0.40	0.09	0.32	7.14	6.55	7.74	**8.88**	4.67	**7.71**	**4**	8.5	8.06	6.57	9.55	4.23	0.92	7.53
GTA V	0.49	0.06	0.43	8.24	7.68	8.80	**7.53**	**6.71**	**10.90**	**1**	**2**	11.73	9.81	13.66	16.90	10.74	23.07
Life is Strange	0.28	0.18	0.10	5.06	4.36	5.76	**6.78**	**8.09**	1.59	**6**	**4.5**	6.20	4.39	8.00	5.63	1.84	9.43
Tom Clancy's The Division	0.40	0.11	0.29	6.81	6.20	7.43	**8.76**	2.97	**8.94**	**5**	**3**	7.41	5.95	8.88	11.97	6.63	17.31
JUMP 'N' RUN	**0.15**	**0.35**	**-0.19**														
Flower	0.11	0.50	-0.39	1.92	1.35	2.49	3.04	**8.20**	1.64	17	14	2.40	1.41	3.39	2.82	0.10	5.54
Journey	0.20	0.22	-0.01	3.66	3.03	4.28	1.00	**5.96**	0.60	22	18.5	1.06	0.48	1.65	0.70	-0.67	2.08
Rayman Legends	0.20	0.26	-0.06	3.55	2.85	4.25	1.76	**9.55**	1.51	11	8.5	3.70	2.35	5.05	4.23	0.92	7.53
Tearaway Unfold	0.09	0.40	-0.31	2.16	1.60	2.72	0.90	**5.31**	0.31	20	16.5	1.66	0.79	2.54	1.41	-0.53	3.35
ROLE-PLAYING	**0.32**	**0.17**	**0.15**														
Dark Souls 3	0.24	0.18	0.06	4.67	4.03	5.31	**9.58**	**5.55**	**7.31**	**2**	**1**	10.5	8.59	12.42	18.31	11.95	24.67
Fallout 4	0.40	0.12	0.28	6.76	6.10	7.42	**9.75**	2.91	**7.61**	**3**	**6**	8.73	7.16	10.30	4.93	1.37	8.49
Final Fantasy X/X – 2 HD Remaster	0.18	0.27	-0.08	3.60	2.99	4.21	4.48	**5.30**	1.17	18	15	2.29	1.34	3.23	2.11	-0.25	4.48
The Witcher 3: Wild Hunt	0.45	0.11	0.34	7.39	6.75	8.01	**6.89**	2.61	3.84	10	12	3.90	2.72	5.08	3.52	0.49	6.55
SHOOTER	**0.22**	**0.23**	**-0.01**														
Call of Duty: Black Ops 3	0.25	0.25	0.00	4.67	3.98	5.36	4.76	2.99	**6.77**	15	21	2.84	1.99	3.70	0.00	0.00	0.00
Uncharted: The Nathan Drake Collection	0.26	0.16	0.10	5.42	4.84	6.00	4.32	1.45	**6.65**	8	12	4.29	2.99	5.60	3.52	0.49	6.55
Wolfenstein: The Old Blood	0.15	0.29	-0.14	2.90	2.37	3.43	2.61	1.35	4.61	21	21	1.37	0.72	2.01	0.00	0.00	0.00
SPORTS	**0.20**	**0.31**	**-0.11**														
NBA 2K16	0.15	0.41	-0.26	2.87	2.21	3.53	0.51	2.33	**5.71**	12.5	**4.5**	3.24	1.83	4.65	5.63	1.84	9.43
Need for Speed	0.28	0.14	0.13	5.37	4.73	6.02	0.45	1.76	**5.93**	16	16.5	2.49	1.47	3.50	1.41	-0.53	3.35
Pro Evolution Soccer	0.18	0.39	-0.21	3.10	2.42	3.77	3.66	5.37	**6.42**	9	18.5	4.18	2.98	5.39	0.70	-0.67	2.08
SURVIVAL HORROR	**0.25**	**0.19**	**0.06**														
Resident Evil – Origins Collection	0.18	0.26	-0.08	3.54	2.97	4.10	**8.09**	3.65	**5.88**	7	12	5.95	4.45	7.44	3.52	0.49	6.55
The Last of Us – Remaster	0.33	0.13	0.20	6.31	5.71	6.91	4.93	1.73	3.07	19	21	1.88	1.01	2.74	0.00	0.00	0.00
FOR FAMILY	**0.14**	**0.49**	**-0.35**														
Guitar Hero Live incl. Guitar	0.17	0.40	-0.23	2.99	2.30	3.67	0.38	**5.81**	0.30	14	8.5	2.89	1.07	4.71	4.23	0.92	7.53
Just Dance 2016	0.11	0.59	-0.47	1.88	1.27	2.49	0.95	**5.71**	1.52	12.5	8.5	3.24	1.93	4.54	4.23	0.92	7.53

Results of open-end question *"Which PS4 game did you miss the most?"*: Maximum of 36 mentions: FIFA 16

As PS4 bundles in the market contain three games at most, the pre-study aimed to identify the best "three-genres-combination" for inclusion in the main study. Eventually, we selected "action-adventure games" (Far Cry Primal, GTA V, Life is Strange, T.C.'s The Division, none), which are important for each of the three preference segments and also best in the TURF analysis, "role-playing games" (Witcher 3, Fallout 4, Final Fantasy X/X2, Dark Souls 3, none) for the same reasoning, and a new genre "games for family and companionship" (Just Dance 2016, Guitar Hero, FIFA 16, none), because the first two games in this category are particularly important for segment two, while FIFA 16 was explicitly missed by 36 out of 142 participants (25.4%) in the open answer question of the study. Other attributes in the main study included hard disk capacity (500GB, 1 TB), color (black, white), the number of controllers (one, two), accessories (charging station, wireless headset, none), and the price (CBC: €299.99, €399.99, €499.99, €599.99, €699.99; ACBC: summed price function).

Experimental design and procedure

The structure of Supplemental Study A corresponded with that of Study 1, by having all participants complete a survey comprising two parts: the conjoint exercise (Part 1) and a HOT (Part 2). However, in this study, we merely disbursed a small group of participants to induce incentive compatibility (see, e.g., Ding et al. 2009; Dong et al. 2010). We gave participants a 1-in-80 chance of winning a PS4 bundle and cash arising from the difference between €700.00 and the respective bundle price. Specifically, we implemented a lottery procedure that rewarded a single participant in each of the three between-subjects conditions: (I) incentive-aligned CBC, (II) hypothetical ACBC, and (III) incentive-aligned ACBC (Figure B2 visualizes the setup).

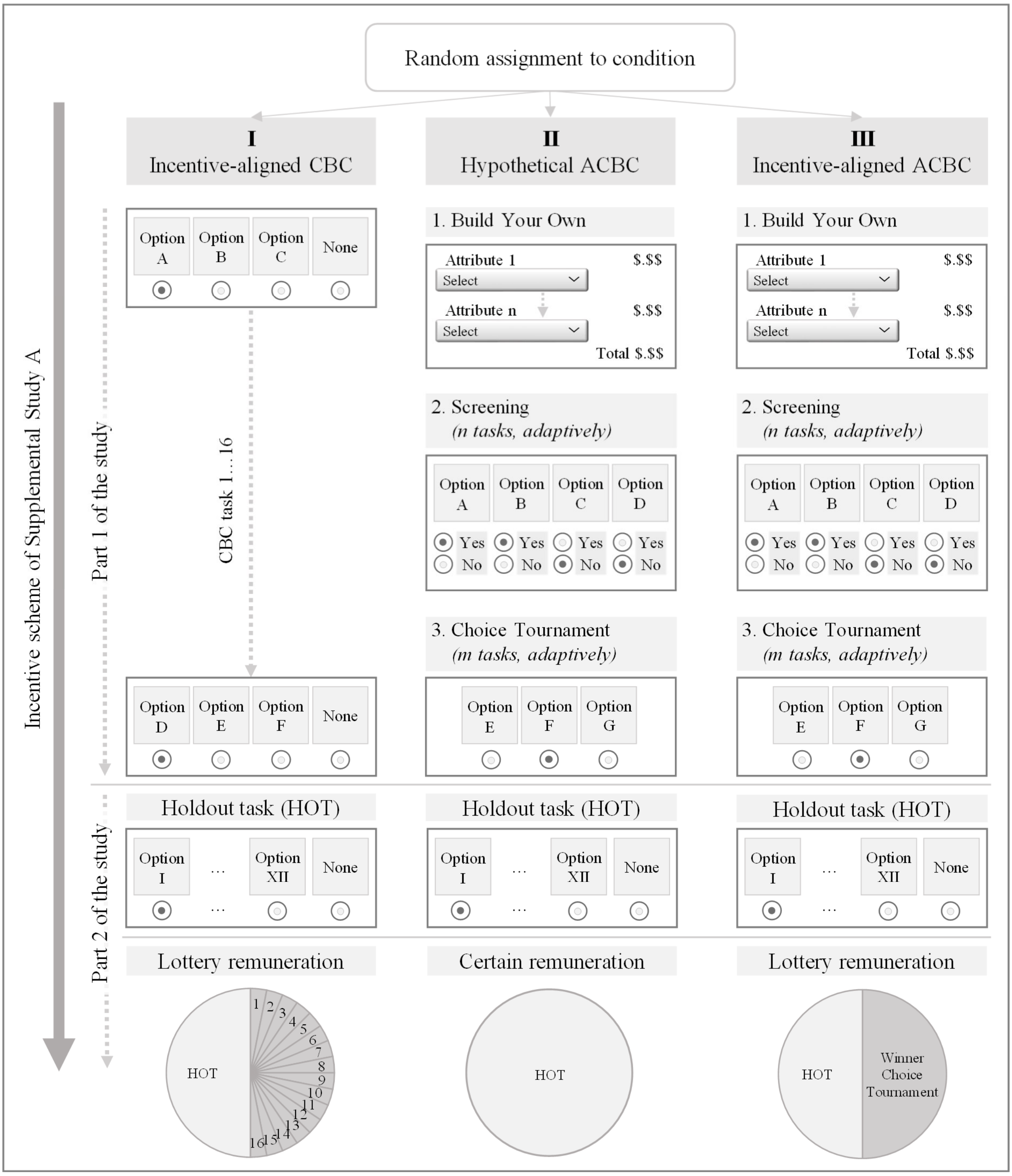


Fig. B2. Incentive scheme of Supplemental Study A

For condition I, we compiled individualized choice designs using a balanced overlap strategy that presented the participants with 16 choice tasks in Part 1, each with three products and a no-purchase option. As in Study 1, we applied the *Direct mechanism* to induce incentive compatibility, rewarding a randomly drawn participant with their own selection in a single randomly drawn CBC task or their own HOT choice. In this case, the HOT included 12 PS4 bundles, presented in a random arrangement, along with the option not to purchase any product (i.e., a free-choice). We optimized an orthogonal design by approaching zero correlations between all attributes, while simultaneously considering all levels in each attribute to appear almost equally often. To enhance ecological validity, we opted for 12 HOT bundles, since Amazon.com usually advertises 12 bundles on the first page of a search request. For conditions II and III, we set up ACBC studies that covered all four ACBC stages, including seven *Screening* tasks with four PS4 bundles each, up to eight choice tasks with three bundles each, and six *Calibration* tasks. For the rest, both conditions followed the procedure of Study 1.

The study took place in the laboratories of the German university. All participants received oral and written instructions on the random lottery mechanism. By means of a manipulation check – requiring participants to select a reward that, according to them, corresponds to a scenario presenting a fictive winning situation – we also tested whether they understood the reward procedure. Participants who did not respond correctly were provided with the right answer and the lottery description before exposure to a second and, if necessary, a third question. We excluded participants who failed after three attempts (12 out of 254).

After the study's completion, we debriefed and thanked all participants. The prize draw took place after conducting the study. To determine the winners in the incentive-aligned conditions I (CBC) and III (ACBC), a coin toss decided which decision (conjoint task or HOT) became payoff relevant.

Participants

All participants (1) had sufficient interest in PS4 bundles, (2) were willing to spend about 30 minutes on a computerized study, (3) were at least 16 years old, (4) were responsible for the PS4-bundle purchase decision in their household, and (5) did not already own a PS4. The net sample consisted of 242 participants, of whom 84 underwent the incentive-aligned CBC (I), 79 the hypothetical ACBC (II), and 79 the incentive-aligned ACBC (III) (38% females, 92% students, 94% up to 30 years of age, 88% with a maximum of €999.99 monthly net income).

The participants' characteristics did not differ across the conditions except for net income (gender: Fisher's exact test, $p = 0.317$; age: Kruskal-Wallis test, $\chi^2_{(2)} = 3.05$; $p = 0.218$; net income: Kruskal-Wallis test, $\chi^2_{(2)} = 8.63$, $p = 0.013$; ownership Sony PlayStation products: Fisher's exact test, $p = 0.619$). We considered the potential influence of monthly net income in our analyses.

Analysis and results

We first compared the hit rates of prediction for three-stage vs. four-stage ACBC to validate the previously reported positive impact of the *Calibration* stage on the predictive validity (e.g., Johnson and Orme 2007). As expected, we found four-stage ACBC to significantly outperform three-stage ACBC (pooled ACBC data: increase in hit rate of 12.03 percentage points, exact McNemar, $p_{(\text{directed})} = 0.001$). As such, our findings justify the integration of the *Calibration* stage into the ACBC procedure. We used data from four-stage ACBC in all analyses. Panel B of Table 2 in the main article provides an overview of the results.

Again, each conjoint approach predicts significantly better in comparison to a prediction by chance (all hit rates > 1/13 = 7.69%, all binomial test *p*-values < 0.001). The concordance of predicted and actual choices (i.e., Cohen's kappa) is *fair* for hypothetical ACBC (0.20) and *moderate* for both incentive-aligned CBC (0.40) and incentive-aligned ACBC (0.42).

Using the same analysis approach as in Study 1, our results once more indicate that incentive-aligned ACBC performs best, with the highest hit rate (56.96%) and MHP (53.33%). It outperforms hypothetical ACBC (hit rate: 44.30%; $\beta = 0.51$, $z = 1.59$, $p_{(\text{one-tailed})} = 0.056$, MHP: 43.43%; $\beta = 0.28$, $t_{(239)} = 1.80$, $p_{(\text{one-tailed})} = 0.037$), as well as incentive-aligned CBC (hit rate: 48.81%; $\beta = 0.33$, $z = 1.04$, $p_{(\text{one-tailed})} = 0.149$, MHP: 35.07%; $\beta = 0.40$, $t_{(239)} = 2.59$, $p_{(\text{one-tailed})} = 0.005$). Again, no significant differences emerge between hypothetical ACBC and incentive-aligned CBC (hit rate: $\beta = -0.18$, $z = -0.58$, $p = 0.565$, MHP: $\beta = 0.12$, $t_{(239)} = 0.77$, $p = 0.444$). Finally, the predictive performance at an aggregate level confirms the superior performance of incentive-aligned ACBC, indicating the lowest mean absolute error (1.52%).

Appendix C: Further information about pre-studies, experimental designs and samples

Preliminary Study to Study 1

Based on an in-depth interview with the management of the partnered pizza restaurant, we gathered information about the product range and common configuration options offered. Then, we conducted a preliminary online study using *Maximum Difference Scaling/ Best-Worst-Scaling* (Louviere et al. 2013; Louviere and Woodworth 1991a) to reduce the restaurant's pizza variety to eight core products most favored by our target sample. In total, 148 pizza consumers (recruited via convenience online sampling) evaluated 18 pizzas in 24 *Best-Worst tasks*, each of which provided five pizzas at once (74% females, M_{age} = 22.80 years [SD_{age} = 4.06], 12% vegetarians, 57% monthly consumers, 76% students). Figure C1 provides a translated screenshot of one of the corresponding MaxDiff tasks. To encourage study participation, we raffled a €20 Amazon voucher among all participants. All of them passed through a screening procedure ensuring that they (1) had sufficient interest in pizza menus, (2) were willing to spend about 10 minutes on an online survey, (3) were at least 16 years old, and (4) were in charge of the purchase decision in their household.

We considered different preference segments to ensure that our final pizza selection satisfied all consumer segments. Specifically, we conducted a Latent Class Analysis in addition to a Hierarchical Bayes (HB) Analysis (Burke et al. 2010; Lagerkvist et al. 2012) to explore preference heterogeneity. The outcome of the Latent Class Analysis, especially the Consistent Akaike Information Criterion (CAIC), suggests a three-cluster solution. Moreover, we also considered the results of HB separately for males and females. Table C1 provides an overview of all pre-study results.

Based on our analyses, we sought to find the combination, which incorporates eight well-favored pizzas, while simultaneously ensuring that the combination provides relevant offers to each identified subgroup (males, females, Latent Class clusters). As a result, pizza types that were not important for any preference segment (e.g., Hot Dog) did not enter Study 1.

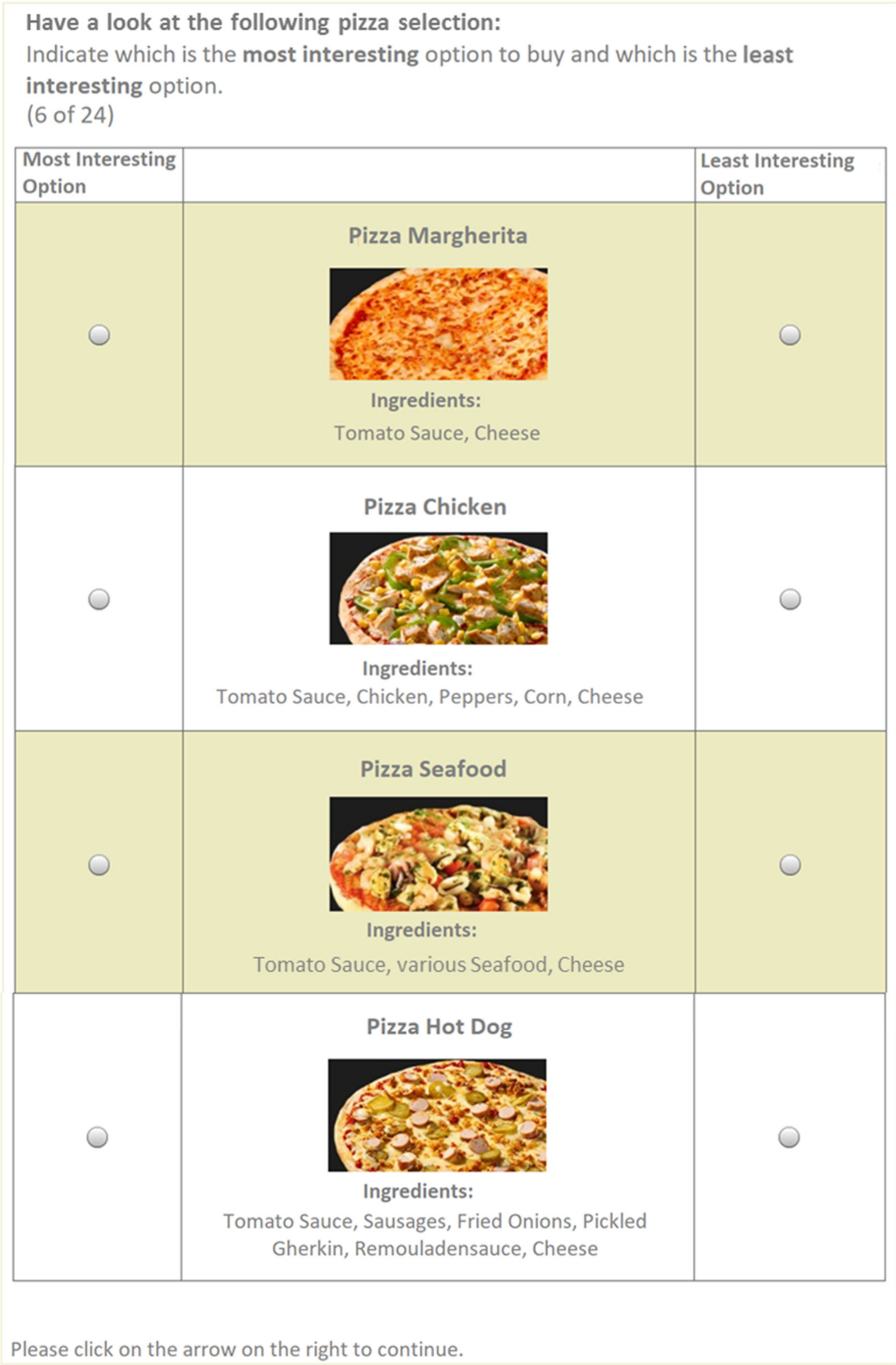


Fig. C1. Exemplary MaxDiff task in the preliminary study of Study 1 (translated)

Table C1. Overview over the main findings in the MaxDiff survey on pizza

***n* = 148**	**MaxDiff counts**			**HB rescaled probability scores**			**HB rescaled probability scores MALES**			**HB rescaled probability scores FEMALES**			**Latent class**		
Pizza type	**Count proportion**				**CI 95%**			**CI 95%**			**CI 95%**		**Solution for 3 groups**		
													I	**II**	**III**
	Best	Worst	Diff	Avg.	Lower	Upper	Avg.	Lower	Upper	Avg.	Lower	Upper	48.3%	20.9%	30.8%
Margherita	0.26	0.20	0.05	**6.26**	5.47	7.05	5.71	4.17	7.25	**6.44**	5.52	7.36	5.65	1.99	**9.12**
Salami	0.24	0.17	0.07	**6.24**	5.47	7.01	6.59	5.03	8.14	**6.08**	5.18	6.97	**9.13**	1.47	4.99
Ham	0.23	0.19	0.04	5.81	5.05	6.56	6.58	5.15	8.02	5.45	4.57	6.33	7.58	2.32	4.65
Salami & Ham	0.26	0.18	0.07	5.82	5.01	6.62	**7.82**	6.18	9.47	5.21	4.29	6.12	**10.57**	1.79	3.46
Funghi	0.29	0.19	**0.10**	**6.39**	5.55	7.23	5.41	3.80	7.02	**6.73**	5.76	7.69	4.56	6.64	**8.59**
Salami, Ham & Mushrooms	0.27	0.17	0.10	5.87	5.07	6.68	**7.63**	6.05	9.21	5.36	4.44	6.28	**8.72**	4.10	3.69
Chicken	0.32	0.16	**0.17**	**6.57**	5.72	7.41	**7.98**	6.24	9.71	**6.17**	5.23	7.11	7.47	**10.04**	4.30
Hot Dog	0.16	0.35	-0.19	3.80	3.00	4.60	5.24	3.59	6.89	3.27	2.41	4.13	4.16	3.22	1.51
Tuna	0.21	0.36	-0.14	4.39	3.51	5.27	4.98	3.21	6.74	4.15	3.15	5.15	3.07	7.73	2.01
Shrimps	0.12	0.48	-0.36	2.59	1.86	3.32	2.95	1.50	4.41	2.49	1.66	3.32	1.15	**10.48**	0.84
Broccoli	0.26	0.22	0.04	**6.17**	5.31	7.03	3.40	1.98	4.81	**7.16**	6.17	8.14	2.85	6.02	**10.67**
4-Cheese	0.31	0.18	**0.12**	**6.61**	5.80	7.43	**7.01**	5.45	8.56	**6.47**	5.52	7.42	6.49	4.48	7.52
Hawaii	0.27	0.22	0.05	5.69	4.83	6.55	5.21	3.64	6.78	5.82	4.81	6.84	6.35	4.70	4.46
Tomato	0.34	0.14	**0.20**	**7.42**	6.63	8.20	5.21	3.59	6.83	**8.19**	7.32	9.05	6.01	3.98	**12.50**
Vegetarian	0.23	0.25	-0.02	5.26	4.40	6.12	3.06	1.72	4.40	5.96	4.96	6.97	2.36	5.69	**9.74**
Spinach	0.30	0.21	0.09	**6.37**	5.51	7.23	4.10	2.62	5.57	**7.11**	6.14	8.09	3.80	6.65	**9.93**
Seafood	0.12	0.60	-0.48	2.46	1.70	3.23	1.85	0.60	3.11	2.66	1.74	3.58	0.54	**10.84**	0.63
Gyros	0.31	0.23	0.09	**6.28**	5.35	7.22	**9.27**	7.59	10.95	5.29	4.24	6.34	**9.53**	7.85	1.39

Overview of incentive scheme in Studies 2 and 3

Studies 2 and 3 follow a structure similar to the previous studies. However, to better understand commonalities and, more importantly, differences (e.g., a validation sample in Study 2 or the *RankOrder* mechanism that requires the utility estimation in Study 3), the following Figures C2 and C3 depict the incentive schemes.

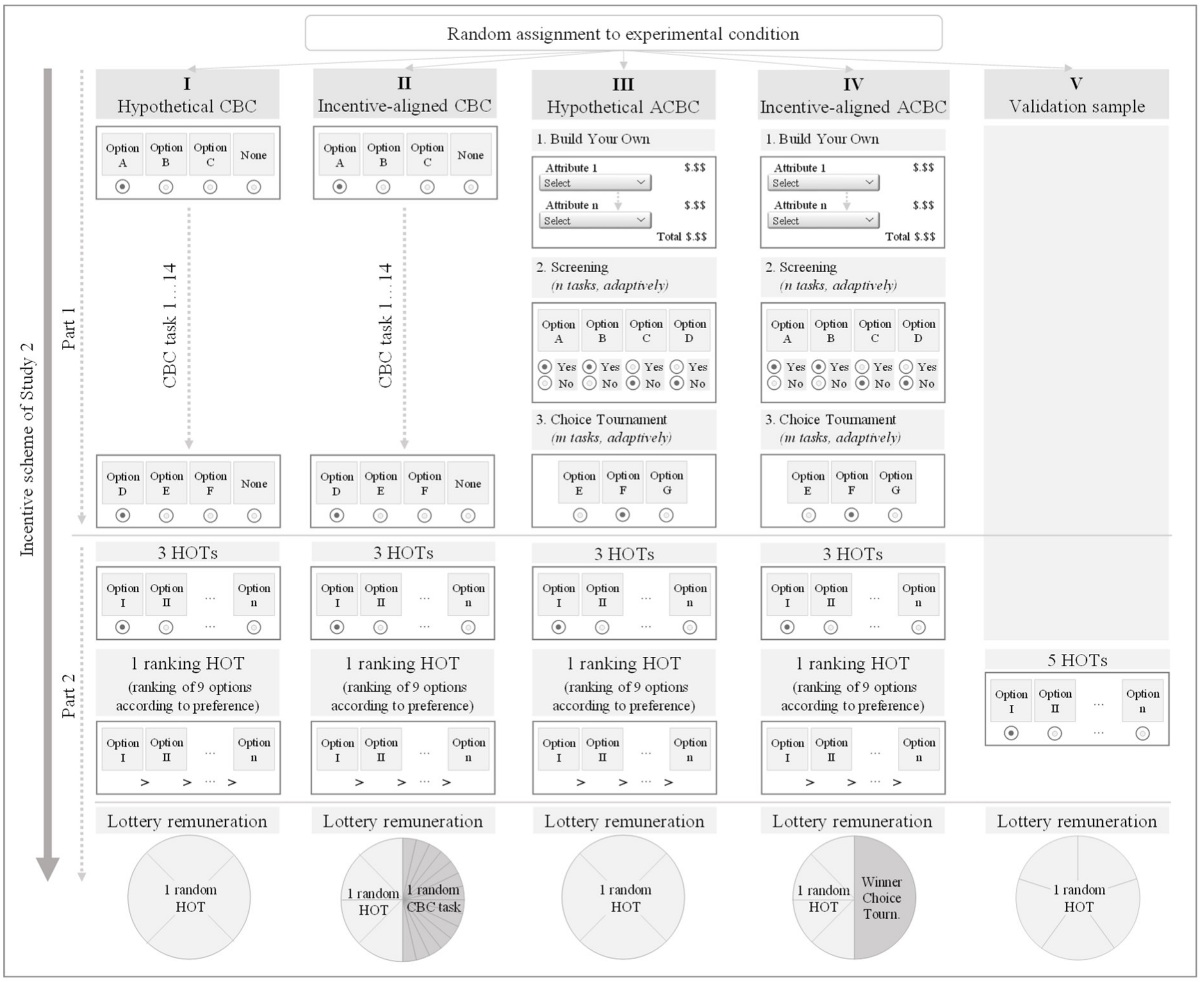


Fig. C2. Incentive scheme of Study 2

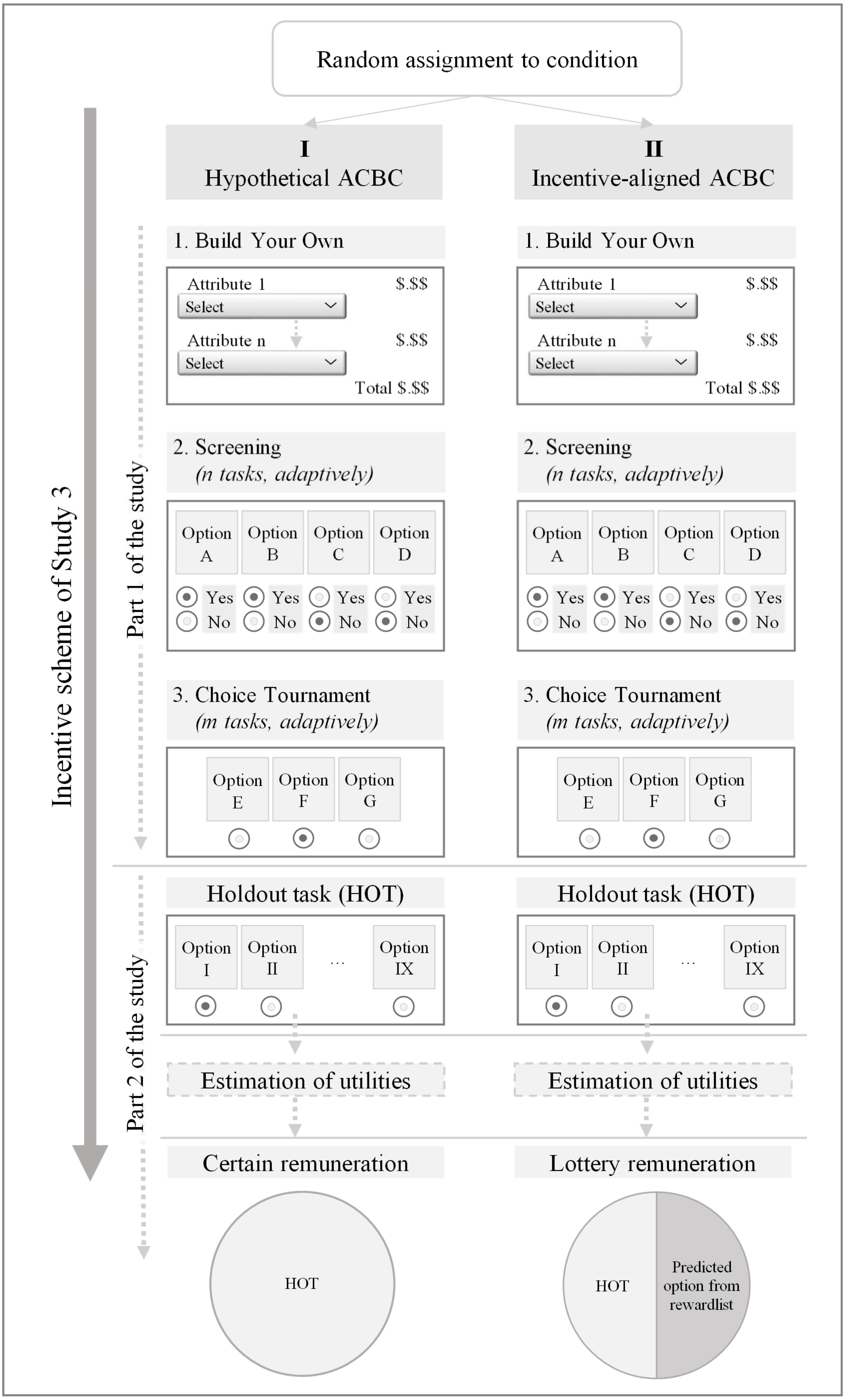


Fig. C3. Incentive scheme of Study 3

Sample characteristics

Table C2. Demographics and pizza-eating behavior: Study 1

		Incentive-aligned CBC	Hypothetical ACBC	Incentive-aligned ACBC
		n=95	n=93	n=90
Gender	female	35 (37%)	51 (55%)	42 (47%)
	male	60 (63%)	42 (45%)	48 (53%)
Age	16-20	11 (12%)	19 (20%)	19 (21%)
	21-25	61 (64%)	54 (58%)	49 (54%)
	26-30	17 (18%)	18 (20%)	14 (16%)
	31 and more	6 (6%)	2 (2%)	8 (9%)
Net income	less than €500	32 (34%)	36 (39%)	25 (28%)
	€500 – €999.99	52 (55%)	50 (54%)	55 (61%)
	€1,000 – €1,499.99	9 (9%)	7 (7%)	6 (7%)
	€1,500 – €1,999.99	1 (1%)	0 (0%)	0 (0%)
	€2,000 and more	0 (0%)	0 (0%)	3 (3%)
	prefer not to say	1 (1%)	0 (0%)	1 (1%)
Pizza consump.	less than 1x / month	18 (19%)	27 (29%)	19 (21%)
	monthly	68 (72%)	57 (61%)	60 (67%)
	weekly	9 (9%)	9 (10%)	11 (12%)
Eating behavior	no constraints	85 (90%)	77 (83%)	74 (82%)
	vegetarian	7 (7%)	13 (14%)	9 (10%)
	vegan	0 (0%)	0 (0%)	0 (0%)
	other	3 (3%)	3 (3%)	7 (8%)

Table C3. Demographics and ownership of Sony products: Supplemental Study A

		Incentive-aligned CBC	Hypothetical ACBC	Incentive-aligned ACBC
		$n = 84$	$n = 79$	$n = 79$
Gender	female	36 (43%)	32 (41%)	25 (32%)
	male	48 (57%)	47 (59%)	54 (68%)
Age	16-20	19 (23%)	14 (18%)	8 (10%)
	21-25	41 (49%)	42 (53%)	44 (56%)
	26-30	23 (27%)	14 (18%)	22 (28%)
	31-35	1 (1%)	5 (6%)	5 (6%)
	36 and more	0 (0%)	4 (5%)	0 (0%)
Net income	less than €500	37 (44%)	27 (34%)	19 (24%)
	€500 – €999.99	39 (46%)	44 (55%)	47 (59%)
	€1,000 – €1,499.99	5 (6%)	2 (3%)	3 (4%)
	€1,500 – €1,999.99	3 (4%)	0 (0%)	3 (4%)
	€2,000 and more	0 (0%)	4 (5%)	7 (9%)
	prefer not to say	0 (0%)	2 (3%)	0 (0%)
Ownership of Sony PS products (not PS4)	yes	49 (58%)	43 (54%)	40 (51%)
	no	35 (42%)	36 (46%)	39 (49%)

Table C4. Demographics and product interest: Study 2

		Hypothetical CBC	Incentive CBC	Hypothetical ACBC	Incentive ACBC	Validation sample
		$n = 105$	$n = 107$	$n = 110$	$n = 100$	$n = 107$
Gender	female	52 (50%)	57 (53%)	68 (62%)	55 (55%)	62 (58%)
	male	53 (50%)	50 (47%)	42 (38%)	45 (45%)	45 (42%)
Occupation	student	29 (27%)	22 (21%)	23 (21%)	24 (24%)	33 (30%)
	employee	60 (57%)	68 (63%)	64 (58%)	52 (52%)	57 (53%)
	freelancer	3 (3%)	4 (4%)	6 (5%)	8 (8%)	6 (6%)
	retiree	7 (7%)	8 (7%)	10 (9%)	8 (8%)	5 (5%)
	other	6 (6%)	5 (5%)	7 (7%)	8 (8%)	6 (6%)
Age	Mean	37.78	37.78	39.47	37.21	35.73
	SD	15.55	14.18	14.16	12.73	12.99
Net income	less than €500	9 (8%)	6 (6%)	11 (10%)	6 (6%)	10 (9%)
	€500 – €999.99	17 (16%)	19 (18%)	19 (17%)	24 (24%)	25 (23%)
	€1,000 – €1,499.99	25 (24%)	10 (9%)	14 (13%)	13 (13%)	18 (17%)
	€1,500 – €1,999.99	19 (18%)	22 (21%)	17 (15%)	17 (17%)	10 (9%)
	€2,000 – €2,499.99	10 (9%)	15 (14%)	17 (15%)	12 (12%)	17 (16%)
	€2,500 – €2,999.99	5 (5%)	9 (8%)	7 (7%)	12 (12%)	8 (8%)
	€3,000 and more	14 (13%)	13 (12%)	14 (13%)	7 (7%)	14 (13%)
	prefer not to say	7 (7%)	13 (12%)	11 (10%)	9 (9%)	5 (5%)
Region	Northern Germany*	29 (27%)	28 (26%)	28 (25%)	21 (21%)	22 (21%)
	Southern Germany*	11 (11%)	17 (16%)	15 (14%)	12 (12%)	8 (8%)
	Western Germany*	14 (13%)	12 (11%)	15 (14%)	18 (18%)	22 (21%)
	Eastern Germany*	51 (49%)	50 (47%)	52 (47%)	49 (49%)	55 (50%)
Interest in food processors	Extraordinary	26 (24%)	15 (14%)	23 (21%)	15 (15%)	22 (21%)
	Considerable	42 (40%)	40 (37%)	48 (44%)	40 (40%)	42 (39%)
	Moderate	25 (24%)	37 (35%)	28 (25%)	39 (39%)	28 (26%)
	Little	12 (12%)	15 (14%)	11 (10%)	6 (6%)	15 (14%)

* Northern Germany: Bremen, Hamburg, Mecklenburg-Western Pomerania, Lower Saxony, Schleswig-Holstein
Southern Germany: Bavaria, Baden-Wuerttemberg
Western Germany: Hessia, North Rhine-Westphalia, Rhineland-Palatinate, Saarland
Eastern Germany: Berlin, Brandenburg, Saxony, Saxony-Anhalt, Thuringia

Table C5. Demographics and product knowledge: Study 3

		Hypothetical ACBC	Incentive-aligned ACBC
		n = 105	*n* = 104
Gender	female	72 (69%)	65 (62%)
	male	33 (31%)	39 (38%)
Occupation	student	87 (83%)	84 (81%)
	employee	13 (12%)	14 (13%)
	trainee	2 (2%)	1 (1%)
	unemployed	3 (3%)	2 (2%)
	other	0 (0%)	3 (3%)
Age	Mean	25.03	26.30
	SD	6.09	7.61
Net income	less than €500	44 (42%)	24 (23%)
	€500 – €1499.99	47 (45%)	55 (53%)
	€1,500 – €2,499.99	9 (9%)	6 (6%)
	€2,500 – €3,499.99	2 (2%)	2 (2%)
	€3,500 and more	0 (0%)	3 (3%)
	prefer not to say	3 (3%)	14 (13%)
	no	36 (46%)	39 (49%)
Product knowledge	Mean	3.40	3.45
	SD	1.25	1.25

Appendix D: Details of the model, estimation, and Stan code, as well as selected results

Model, estimation, and Stan code

We employ the same hierarchical choice model for each study and condition. In this Web Appendix, we report details on the particular model specification and estimation procedure, including our Stan code. We also provide summary tables for population means and the standard deviations of each estimation.

Model Participant i $(i = 1, \dots, I)$ has T_i choice tasks and selects from choice task t the alternative j from the available J_t alternatives, which provide the highest utility u_{ijt}. The error-term ε_{ijt} is type I extreme-value distributed with shape value 0 and scale value 1, leading to the multinomial logit (MNL) model at the individual level (Allenby et al. 2014).

$$u_{ijt} = \boldsymbol{x}'_{ijt} \cdot \boldsymbol{\beta_i} - \beta_i^p \cdot price_{ijt} + \varepsilon_{ijt}, \;\; \varepsilon_{ijt} \sim EV(0,1). \qquad \text{(formula D1)}$$

The vector $\boldsymbol{x}_{ijt}$ contains information about the alternative j in choice tasks t as seen by participant i. We use effects coding for all attributes except for the intercept of the "none" option, which is included as a dummy in $\boldsymbol{x}$ and for the price ($price_{ijt}$). For the latter, we employ linear coding and rescale and de-mean raw prices to ensure that the resulting parameters have reasonable magnitudes and are easy to interpret (e.g., in Study 1, we rescale the price in €10, and in the other three studies in €100). De-meaning the price in the model, in combination with the effects coding of the attributes, facilitates the interpretation of the none parameter as it is measured against an "average" product.

$$l(\boldsymbol{y}_i | \boldsymbol{\beta}_i, \beta_i^p) = \prod_t^{T_i} \frac{\exp(\boldsymbol{x}'_{iy_{it}t} \cdot \boldsymbol{\beta}_i - \beta_i^p \cdot price_{iy_{it}t})}{\sum_{j'}^{J_t} \exp(\boldsymbol{x}'_{ij't} \cdot \boldsymbol{\beta}_i - \beta_i^p \cdot price_{ij't})} \qquad \text{(formula D2)}$$

Let $\boldsymbol{y_i} = [y_{i1}, \dots, y_{iT_i}]'$ denote the vector of observed choices of participant i. Hence $l\left(\boldsymbol{y}_i \middle| \boldsymbol{\beta}_i, \beta_i^p\right)$ in formula, D2 is the likelihood of participant i's choice sequence, conditional on the participant-level parameters $\boldsymbol{\beta}_i$ and β_i^p.

While T and J have the same dimension for each participant in regular CBC studies, this is not the case in ACBC studies with adaptive designs, as explicitly highlighted in formula D2. T_i is participant-specific and the number of alternatives in the denominator of the logit probability (J_t) varies over the choice tasks. J_t varies in the *BYO* stage between two and the maximum number of attribute levels. Furthermore, it is two in the *Screening* stage and three in the *Choice Tournament* stage (see Sawtooth Software (2014), for more details).[2]

To account for heterogeneity across participants, as indicated above, all parameters are participant-specific. In particular, $\boldsymbol{\beta}_i$ and β_i^p follow a joint multivariate normal distribution with a full covariance matrix $\boldsymbol{\Sigma}$:

$$\begin{bmatrix} \boldsymbol{\beta}_i \\ \beta_i^{p*} = \ln(\beta_i^p) \end{bmatrix} \sim MVN\left(\begin{bmatrix} \boldsymbol{b} \\ b^p \end{bmatrix}, \boldsymbol{\Sigma}\right) \text{ and} \qquad \text{(formula D3)}$$

$$\boldsymbol{\Sigma} = \operatorname{diag}\left(\begin{bmatrix} \boldsymbol{\sigma} \\ \sigma^p \end{bmatrix}\right) \cdot \boldsymbol{\Omega} \cdot \operatorname{diag}\left(\begin{bmatrix} \boldsymbol{\sigma} \\ \sigma^p \end{bmatrix}\right). \qquad \text{(formula D4)}$$

The utility function (formula D1) contains K parameters ($K - 1$ in $\boldsymbol{\beta}_i$ and one in β_i^p). Hence the dimension of the covariance $\boldsymbol{\Sigma}$ and correlation $\boldsymbol{\Omega}$ matrices is $K \times K$. We separate the price parameter from the other parameters to impose a sign constraint via the lognormal distribution at the individual level and to use different priors.

[2] The inclusion of scale differences over the ACBC stages, as suggested by Otter (2007), is straightforward (i.e., the adaption of the Stan code is simple, and Hamiltonian Monte Carlo does not require an additional layer of Gibbs sampling, as does regular MCMC for the HB-MNL model when some parameters are fixed (Train 2009, chap. 11)). However, we do not include this model feature because our sample sizes are considerably smaller than the recommended lower threshold of 300 participants to achieve stable estimates in Sawtooth Software (2014, p. 16).

To be specific, as $\ln(\beta_i^p)$ follows a normal distribution, all β_i^p are positive, and as we define prices with a minus sign in the utility function, we restrict all participants to having a negative price sensitivity.[3] In Web Appendix G, we explore the consequences of the sign constraint.

The separate specification of the standard deviations ($\boldsymbol{\sigma}$ and σ^p) and the correlation matrix ($\boldsymbol{\Omega}$) facilitates the interpretation of the results and allows us to use certain priors (see the discussion below) that have favorable properties (Barnard et al. 2000). Since we use Bayesian methods for estimation, we need specific (hyper-)priors to complete our model specification:

$$\begin{aligned} \boldsymbol{b} &\sim N(0,5), \\ b^p &\sim N(1,1), \\ \boldsymbol{\sigma} &\sim N_+(0,2.5), \\ \sigma^p &\sim N_+(0,1), \text{ and} \\ \boldsymbol{\Omega} &\sim LKJCorr(2). \end{aligned} \qquad \text{(formulas D5)}$$

For $\boldsymbol{b}$ and $\boldsymbol{\sigma}$ the priors in formulas C5 refer to each element in the vectors. The specifications for the priors of b^p and σ^p ensure that the lognormal distribution does not "pile up" probability mass very close to zero (implying almost no price sensitivity), and limit the already long tail of the distribution to meaningful values. Thus, these specific priors are less informative, compared to ad hoc-chosen priors where the mean is set to zero and/or the standard deviations are larger (see Allenby et al. 2014 for a related discussion).

N_+ denotes the half-normal distribution (i.e., a folded "at zero" normal distribution), and the implied mean values for $\boldsymbol{\sigma}$ and σ^p are $2.5 \cdot \sqrt{2/\pi} = 1.995$ and $1 \cdot \sqrt{2/\pi} = 0.798$, respectively. Please note that the mode of half-normal distribution is by definition zero.

---

[3] Restricting the price parameter to a negative value is useful to ensure that the results do not violate standard economic assumptions (i.e., participants having downward-sloping demands), which could arise because of the small amount of information at the participant level. This option is also available in Sawtooth, even though – in the latter case – the sign constraint is not implemented via the (negative) lognormal distribution, but by "simultaneous tying" (Johnson 2000). Based on our experience, the results are, in most cases, near identical. We opted for the version with the lognormal distribution because it is also used in modeling literature (e.g., Allenby et al. 2014) and is easy to implement. Restricting the price parameter (distribution) to a negative value also guarantees finite moments for the (implied) WTP of the attributes (Daly et al. 2012). This is crucial for our analyses in Web Appendices D and H.

The value of 2 for the *LKJ* prior (Lewandowski et al. 2009) slightly shrinks correlations in $\mathbf{\Omega}$ towards zero. It is useful to have separate priors for the vector of standard deviations and the correlation matrix since this ensures no correlation (of draws) between the standard deviations and the correlations in the covariance matrix, which is not the case for the common (inverse) Wishart prior (see, e.g., Akinc and Vandebroek (2018) for more details of this “separation strategy”). This setup (or near-similar setups) for the prior of the covariance matrix has gained popularity in the statistical literature and is also used by other quantitative marketing researchers (see, e.g., Bruno et al. 2018 or Ellickson et al. 2019).

In summary, given the scaling of prices in each dataset and the effects coding for all attributes, our priors are reasonable but weakly informative. We prefer using weakly informative (but proper) priors because non-informative priors can cause serious problems (Gelman et al. 2014). We did not find our prior choice to have a noticeable effect on our results (also see the case study in Web Appendix E, where we compare the results using our own code with the results from Sawtooth Software’s Lighthouse Studio v. 9.8.1).

Estimation We used Bayesian methods (Train 2009) to estimate the hierarchical MNL model. We specifically implemented the model presented in this Web Appendix in Stan (Carpenter et al. 2017; Stan Development Team 2020a): a probabilistic programming language that enables full Bayesian statistical inference using Hamiltonian Monte Carlo (HMC).

HMC is highly efficient and converges faster for high-dimensional posterior distributions than regular MCMC algorithms (e.g., random walk Metropolis or Gibbs sampling) because it utilizes first-order gradient information (Gelman et al. 2014).

Furthermore, HMC does not require conjugate priors and thus provides more flexibility (i.e., using half-normal priors for standard deviations or having separate priors for correlations and standard deviations). We employed the No-U-turn sampler that automatically tunes the step size of the algorithm to draw samples from the posterior distribution (Hoffman and Gelman 2014).

All results in the paper and this Web Appendix are based on 5,000 draws from the posterior distributions. For each model (i.e., each study and condition), we ran five chains with random starting values for 5,500 iterations, where the first 500 draws were discarded as "warm-up" draws, and every fifth draw was retained for the posterior inference. We used the Gelman-Rubin statistic (Gelman and Rubin 1992) as a formal test of convergence, and the values for all parameters were close to 1 (maximum below 1.004), indicating good convergence (we report specific values in Tables D1 – D12).

A visual inspection of trace plots for population-level parameters and the log-likelihoods in each chain, additionally indicated the desired shape of "fat hairy caterpillars" and thus confirmed convergence, good mixing, and low autocorrelation. As such, we are confident that the HMC algorithm worked well and that the posterior results are reliable. Thus, we combined the draws of the five chains with the 5,000 draws from the posterior distribution (with a median (minimum) effective sample size over all parameters in each estimation of well above 4,000 (2,000)).

Stan code Listing D1 shows the Stan code (`hbmnl_sc.stan`) used for estimation with remarks and comments. Stan code has a specific structure with multiple blocks (data, (transformed) parameters, model, generated quantities). The model notation follows the above-mentioned definitions.

Listing D1. Stan Code for the HB MNL model with sign constraint

```
data {
  int<lower=1> N;  // number of choices (and choice tasks)
  int<lower=1> I;  // number of individuals
  int<lower=1> M;  // number of rows in X
  int<lower=1> K;  // number of cols in X (incl. price in last col)
  int<lower=1> D;  // number of cols in Z
  matrix[M, K] X;  // design matrix
  matrix[I, D] Z;  // individual level variables (e.g., demographics)
  int<lower=1> y[N];  // row number in X that belongs to nth choice task
  int<lower=1> start[N];  // row number in X where nth choice task starts
  int<lower=1> end[N];  // row number in X where nth choice task ends
  int<lower=1, upper=I> id[N];  // id identifying each individual
}

parameters {
  matrix[K, I] z;  // raw heterogeneity
  matrix[D, K] b;  // mean (heterogeneity)
  cholesky_factor_corr[K] L_Omega;  // cholesky factor (correlation)
  vector<lower=0>[K] sigma;  // sd (heterogeneity)
}

transformed parameters {
  vector[M] v;  // deterministic part of the utility
  vector[N] ll;  // log-likelihood values for each observation
  matrix[I, K] beta;  // individual level parameters

  // non-centered parameterization
  beta = Z * b + (diag_pre_multiply(sigma, L_Omega) * z)';
  beta[, K] = -exp(beta[, K]);  // sign-constrained price coef. (negative
lognormal)

  for (n in 1:N) {
    // the start and end indices allow for varying choice set sizes
    v[start[n]:end[n]] = X[start[n]:end[n],] * beta[id[n]]';
    ll[n] = v[y[n]] - log_sum_exp(v[start[n]:end[n]]);  // loglik for each chosen
alt.
  }
}

model {
  to_vector(z) ~ std_normal();  // non-centered parameterization
  L_Omega ~ lkj_corr_cholesky(2);  // slight shrinkage towards 0

  // prior for means and sds of all parameters except price parameter
  to_vector(b[, :(K-1)]) ~ normal(0, 5);
  sigma[:(K-1)] ~ normal(0, 2.5); // implies an avg. sd of 2.5 * sqrt(2/pi) = 1.995

  // different priors for mean and sd of the price coef. (because of log-scale)
  b[, K] ~ normal(1, 1);
  sigma[K] ~ normal(0, 1);  // implies an avg. sd of 1 * sqrt(2/pi) = 0.798

  for (n in 1:N) {
    target += ll[n];  // target also includes the (log densities of the) priors
  }
}

generated quantities {
  matrix[K, K] Omega;
  real<upper=0> log_lik;

  Omega = multiply_lower_tri_self_transpose(L_Omega); // correlation matrix
  log_lik = sum(ll); // log-likelihood value for all obs.
}
```

Several points are worth mentioning to facilitate the understanding of the code:

1. The code allows the inclusion of participant-level information (e.g., demographics) via matrix $\boldsymbol{Z}$. Also, the mean of the population-level parameters ($\boldsymbol{b}$) is a matrix with corresponding dimensions. We do not use this in our applications and hence $\boldsymbol{Z}$ is a $I \times 1$ "matrix" containing only 1s.
2. In most cases, it is more efficient to use (what is known as) non-centered parametrization for hierarchical models in Stan (Stan Development Team (2018), chap. 1.13). Instead of directly sampling from a multivariate normal distribution $\boldsymbol{\beta_i} \sim MVN(\boldsymbol{b}, \boldsymbol{\Sigma})$, we sample a vector $\boldsymbol{z_i}$ from multiple standard normal distributions and transform these draws by using the lower triangular matrix $\boldsymbol{L}$ of the Cholesky decomposition of the covariance matrix $\boldsymbol{\Sigma}$ and adding the mean vector $\boldsymbol{b}$ (i.e., $\boldsymbol{\beta_i} = \boldsymbol{b} + \boldsymbol{L} \cdot \boldsymbol{z_i}$). As we separately model the correlation matrix $\boldsymbol{\Omega}$ and the vector of standard deviations $\boldsymbol{\sigma}$, we have $\boldsymbol{L} = \text{diag}(\boldsymbol{\sigma}) \cdot \boldsymbol{L_\Omega}$, where $\boldsymbol{L_\Omega}$ is the lower triangular matrix of the Cholesky decomposition of the correlation matrix $\boldsymbol{\Omega}$.
3. To obtain the negative lognormal distribution for the price parameter, we simply transform the raw draw β_i^{p*} from the normal distribution via $\beta_i^p = -\exp\,(\beta_i^{p*})$.
4. To deal with each participant's varying number of choice tasks and the varying number of alternatives in different choice tasks (i.e., the "ragged" data structure), we loop over all *N* choice observations and use indices for the corresponding start and end positions of the relevant choice task information in the dataset.
5. As Stan needs the log-density of the model's given data (and automatically adds the log-density of the priors), we directly specify the *log*-likelihood for each observation as where $v_n = \boldsymbol{x}_n' \cdot \boldsymbol{\beta}_n - \beta_n^p \cdot price_n$ is the deterministic part of the utility of the chosen alternative in observation n and v_{nj} is the deterministic utility for the *j*-th alternative in the corresponding choice task.

$$ll_n = v_n - \log\left(\sum_j^{J_n} \exp(v_{nj})\right), \qquad \text{(formula D6)}$$

In this case, we also use the index n for the parameter vector because each observation belongs to a specific participant i.

For specific details of the Stan language, please refer to the manual and functions reference (Stan Development Team 2020b). We saved the code in Listing D1 as a .stan file, enabling it to be called through R using the package RStan (Stan Development Team 2020a). However, interfaces are available for Python, Matlab, Julia, etc. For more information on our code, the interested reader is referred to the replication files, which are available within the Open Science Framework (https://osf.io/e5v4c/). The case study in Web Appendix E provides more information on the R code to load and reshape data, as well as on the model estimation (including the *Calibration* stage).

The next three subsections briefly summarize the *estimation results* for all studies. We start with population-level estimates, continue with partworths as well as relative importance values and finally provide an overview regarding WTP. Each part provides important additional information to better understand differences and commonalities across conditions (within each study).

Population-level estimates

Tables D1 to D12 summarize the estimation results for each condition in each study. We report posterior means and the 95% credible intervals for $\boldsymbol{b}$ and $\boldsymbol{\sigma}$ (i.e., the means and the standard deviations of all parameters in the population). For the price parameter, we report the results in the main part of the table before applying the exponential transformation to the normal distribution because of the sign constraint. The implied average price sensitivity and the corresponding standard deviation in the population at the mean of the results can be easily computed from the posterior results using closed-form transformations based on the log-normal distribution:

$$\hat{b}^p = -\exp(\hat{b}^{p*} + \hat{\sigma}^{p*2}/2) \text{ and} \qquad \text{(formula D7)}$$

$$\hat{\sigma}^p = \sqrt{\left(\exp\left(\hat{\sigma}^{p*2}\right) - 1\right) \cdot \exp(2 \cdot \hat{b}^{p*} + \hat{\sigma}^{p*2}).} \qquad \text{(formula D8)}$$

The transformed results can be found in the notes of each table.

We generally find the following: (1) The Bayesian estimation worked and converged well (i.e., high number of effective sample sizes and max($\hat{\mathrm{R}}$)-value close to 1). (2) Most values for $\hat{\boldsymbol{b}}$ are face-valid (e.g., participant prefer on average larger pizza over smaller pizza in Study 1 or food processors with more power in Study 2) and statistically in the sense that the 95%-credible interval does not cover zero. (3) Preference heterogeneity plays a crucial role and is statistically significant, but the values for $\boldsymbol{\sigma}$ are still reasonable in magnitude. (4) By tendency, we find stronger price effects in conditions with incentive alignment and/or adaptive designs. Note, however, that direct comparison of utility parameters across datasets is difficult. We address this point also in other parts of this Web Appendix, where we discuss the (effect of the) sign constraint of the price parameter (Web Appendix G), the scale of the models (Web Appendix H), and the price elasticity (Web Appendix H). Overall, we conclude that all estimations provided robust and useful results.

Table D1. Summary of estimation results for Study 1: Incentive-aligned CBC

	$\widehat{\boldsymbol{b}}$ and $\hat{b}^{p*}$		$\widehat{\boldsymbol{\sigma}}$ and $\hat{\sigma}^{p*}$	
Parameter	**Mean**	**95% CI**	**Mean**	**95% CI**
None	3.389	[2.499, 4.373]	3.637	[2.867, 4.527]
Margherita	-0.921	[-1.724, -0.229]	1.891	[1.211, 2.680]
Salami	0.494	[-0.084, 1.062]	1.671	[1.047, 2.403]
Fungi	-0.228	[-0.984, 0.450]	2.207	[1.496, 3.034]
Salami, ham & fungi	-0.146	[-0.852, 0.487]	1.951	[1.296, 2.713]
Chicken	0.354	[-0.300, 0.964]	1.906	[1.246, 2.671]
Four cheeses	0.176	[-0.498, 0.797]	2.021	[1.387, 2.800]
Tomato	0.618	[-0.024, 1.248]	1.750	[1.125, 2.487]
20cm	-0.604	[-0.930, -0.293]	0.747	[0.347, 1.155]
25cm	0.047	[-0.193, 0.297]	0.257	[0.012, 0.656]
Sauce hollandaise	-0.456	[-0.803, -0.113]	1.110	[0.762, 1.513]
BBQ sauce	-0.065	[-0.367, 0.225]	0.815	[0.472, 1.194]
Onion rings	0.546	[0.144, 0.943]	0.780	[0.116, 1.405]
Paprika	0.384	[0.012, 0.751]	0.591	[0.037, 1.243]
Jalapenos	0.368	[-0.086, 0.802]	1.158	[0.651, 1.699]
Olives	-0.357	[-0.870, 0.113]	1.389	[0.853, 2.009]
Boiled egg	-1.417	[-2.079, -0.836]	1.523	[0.916, 2.225]
Gouda	0.226	[-0.037, 0.496]	0.254	[0.013, 0.650]
Feta	-0.087	[-0.401, 0.222]	0.711	[0.229, 1.150]
Mozzarella	0.291	[-0.009, 0.578]	0.607	[0.151, 1.036]
Garlic	0.259	[0.068, 0.456]	0.399	[0.084, 0.690]
Coca-ola	0.426	[0.076, 0.794]	0.638	[0.083, 1.161]
Coca-Cola light or zero	0.468	[0.119, 0.831]	0.415	[0.021, 0.971]
Fanta	-0.236	[-0.603, 0.113]	0.335	[0.012, 0.871]
Sprite	-0.141	[-0.518, 0.213]	0.487	[0.026, 1.129]
Water (still or medium)	-0.111	[-0.472, 0.236]	0.406	[0.021, 0.980]
Price	1.505	[1.134, 1.818]	0.852	[0.608, 1.163]

Notes: $\mathrm{LMD} = -876.300$, $\min(\mathrm{n_{eff}}) = 2714$, $\mathrm{median}(\mathrm{n_{eff}}) = 4811$, and $\max(\widehat{\mathrm{R}}) = 1.002$. The implied average price sensitivity and the corresponding standard deviation in the population at the mean of the results are $\hat{b}^{p} = -6.475$ and $\hat{\sigma}^{p} = 6.687$, respectively.

Table D2. Summary of estimation results for Study 1: Hypothetical ACBC

	$\hat{\boldsymbol{b}}$ and $\hat{b}^{p*}$		$\hat{\boldsymbol{\sigma}}$ and $\hat{\sigma}^{p*}$	
Parameter	**Mean**	**95% CI**	**Mean**	**95% CI**
None	3.778	[3.197, 4.386]	1.802	[1.394, 2.266]
Margherita	-0.666	[-1.236, -0.146]	2.153	[1.704, 2.668]
Salami	-0.042	[-0.663, 0.555]	2.578	[2.082, 3.159]
Fungi	-0.724	[-1.432, -0.065]	2.707	[2.144, 3.367]
Salami, ham & fungi	0.663	[0.086, 1.202]	2.093	[1.655, 2.609]
Chicken	-0.241	[-0.838, 0.302]	2.144	[1.645, 2.735]
Four cheeses	0.891	[0.414, 1.357]	1.709	[1.304, 2.177]
Tomato	0.730	[0.171, 1.279]	2.113	[1.670, 2.630]
20cm	-1.421	[-1.781, -1.065]	1.320	[0.991, 1.692]
25cm	0.478	[0.315, 0.647]	0.336	[0.027, 0.638]
Sauce hollandaise	-0.892	[-1.340, -0.482]	1.589	[1.230, 1.993]
BBQ sauce	-0.720	[-1.112, -0.354]	1.278	[0.933, 1.659]
Onion rings	0.165	[-0.285, 0.617]	1.594	[1.195, 2.060]
Paprika	0.682	[0.293, 1.050]	1.186	[0.809, 1.588]
Jalapenos	0.337	[-0.153, 0.797]	1.745	[1.325, 2.241]
Olives	-0.779	[-1.367, -0.248]	2.035	[1.535, 2.583]
Boiled egg	-1.499	[-2.173, -0.878]	2.293	[1.739, 2.932]
Gouda	-0.308	[-0.545, -0.083]	0.372	[0.032, 0.704]
Feta	-0.025	[-0.285, 0.225]	0.720	[0.464, 0.994]
Mozzarella	0.334	[0.122, 0.542]	0.389	[0.040, 0.721]
garlic	-0.063	[-0.204, 0.073]	0.446	[0.264, 0.639]
Coca-cola	0.513	[0.185, 0.822]	1.020	[0.688, 1.397]
Coke light or zero	-0.447	[-0.884, -0.038]	1.519	[1.130, 1.951]
Fanta	0.051	[-0.259, 0.348]	0.614	[0.090, 1.060]
Sprite	0.051	[-0.288, 0.368]	0.820	[0.402, 1.217]
Water (still or medium)	0.196	[-0.168, 0.535]	1.149	[0.810, 1.515]
Price	1.714	[1.498, 1.909]	0.633	[0.470, 0.821]

Notes: $\text{LMD} = -2257.060$, $\min(\text{n}_{\text{eff}}) = 2047$, $\text{median}(\text{n}_{\text{eff}}) = 4597$, and $\max(\hat{\text{R}}) = 1.002$. The implied average price sensitivity and the corresponding standard deviation in the population at the mean of the results are $\hat{b}^{p} = -6.782$ and $\hat{\sigma}^{p} = 4.762$, respectively.

Table D3. Summary of estimation results for Study 1: Incentive-aligned ACBC

	$\hat{\boldsymbol{b}}$ and $\hat{b}^{p*}$		$\hat{\boldsymbol{\sigma}}$ and $\hat{\sigma}^{p*}$	
Parameter	**Mean**	**95% CI**	**Mean**	**95% CI**
None	4.574	[3.842, 5.376]	2.698	[2.197, 3.285]
Margherita	-0.238	[-0.787, 0.291]	2.005	[1.499, 2.592]
Salami	0.310	[-0.210, 0.810]	1.814	[1.367, 2.338]
Fungi	-1.328	[-2.122, -0.562]	2.920	[2.260, 3.713]
Salami, ham & fungi	0.100	[-0.650, 0.779]	2.565	[1.997, 3.245]
Chicken	0.635	[-0.056, 1.281]	2.449	[1.901, 3.068]
Four cheeses	0.618	[0.116, 1.091]	1.436	[0.958, 1.962]
Tomato	0.734	[0.196, 1.254]	1.928	[1.439, 2.465]
20cm	-0.914	[-1.261, -0.583]	1.182	[0.842, 1.560]
25cm	0.420	[0.233, 0.605]	0.320	[0.023, 0.651]
Sauce hollandaise	-0.711	[-1.149, -0.291]	1.480	[1.119, 1.900]
BBQ sauce	-0.612	[-1.007, -0.244]	1.310	[0.969, 1.702]
Onion rings	-0.174	[-0.699, 0.312]	1.619	[1.171, 2.121]
Paprika	0.846	[0.525, 1.178]	0.612	[0.106, 1.060]
Jalapenos	0.284	[-0.206, 0.762]	1.579	[1.162, 2.076]
Olives	-0.869	[-1.442, -0.329]	1.745	[1.252, 2.318]
Boiled egg	-1.469	[-2.165, -0.847]	2.124	[1.621, 2.728]
Gouda	-0.024	[-0.301, 0.230]	0.492	[0.165, 0.797]
Feta	-0.446	[-0.725, -0.197]	0.400	[0.040, 0.764]
Mozzarella	0.369	[0.154, 0.582]	0.224	[0.011, 0.542]
garlic	-0.076	[-0.230, 0.068]	0.469	[0.284, 0.660]
Coca-Cola	0.010	[-0.352, 0.335]	0.989	[0.635, 1.380]
Coke light or zero	-0.088	[-0.441, 0.234]	0.745	[0.317, 1.161]
Fanta	0.120	[-0.180, 0.395]	0.456	[0.046, 0.855]
Sprite	-0.194	[-0.525, 0.113]	0.488	[0.043, 0.948]
Water (still or medium)	0.077	[-0.245, 0.377]	0.806	[0.432, 1.173]
Price	1.904	[1.672, 2.115]	0.782	[0.606, 1.003]

Notes: $LMD = -2081.327$, $\min(n_{eff}) = 2767$, $\text{median}(n_{eff}) = 4701$, and $\max(\hat{R}) = 1.002$. The implied average price sensitivity and the corresponding standard deviation in the population at the mean of the results are $\hat{b}^p = -9.114$ and $\hat{\sigma}^p = 8.369$, respectively.

Table D4. Summary of estimation results for Supplemental Study A: Incentive-aligned CBC

	$\hat{\boldsymbol{b}}$ and $\hat{b}^{p*}$		$\hat{\boldsymbol{\sigma}}$ and $\hat{\sigma}^{p*}$	
Parameter	**Mean**	**95% CI**	**Mean**	**95% CI**
None	3.024	[2.362, 3.774]	2.262	[1.749, 2.878]
500GB	-0.447	[-0.657, -0.250]	0.530	[0.229, 0.838]
Black	-0.041	[-0.237, 0.151]	0.516	[0.280, 0.775]
One dual shock controller	-0.430	[-0.679, -0.203]	0.768	[0.532, 1.047]
Charging station	0.160	[-0.053, 0.366]	0.179	[0.007, 0.490]
Wireless headset	0.141	[-0.100, 0.383]	0.533	[0.130, 0.893]
Far Cry Primal	0.390	[0.021, 0.748]	0.760	[0.143, 1.335]
GTA V	0.880	[0.454, 1.319]	1.288	[0.848, 1.773]
Life is Strange	0.043	[-0.362, 0.435]	1.096	[0.537, 1.666]
TC's The Division	-0.115	[-0.568, 0.312]	1.255	[0.758, 1.817]
Witcher 3	0.353	[-0.081, 0.792]	1.297	[0.789, 1.883]
Fallout 4	0.537	[0.204, 0.866]	0.559	[0.042, 1.110]
Final Fantasy x/x2	-0.330	[-0.743, 0.057]	1.032	[0.532, 1.556]
Dark Souls 3	0.175	[-0.206, 0.555]	0.826	[0.193, 1.383]
Just Dance 2016	-0.500	[-0.972, -0.061]	1.504	[1.109, 1.997]
Guitar Hero	-0.048	[-0.495, 0.373]	1.433	[0.973, 1.974]
FIFA 16	0.723	[0.263, 1.197]	1.609	[1.195, 2.113]
Price	0.345	[0.133, 0.555]	0.620	[0.474, 0.791]

Notes: $\text{LMD} = -797.798$, $\min(\text{n}_{\text{eff}}) = 3162$, $\text{median}(\text{n}_{\text{eff}}) = 4684$, and $\max(\hat{\text{R}}) = 1.003$. The implied average price sensitivity and the corresponding standard deviation in the population at the mean of the results are $\hat{b}^p = -1.711$ and $\hat{\sigma}^p = 1.172$, respectively.

Table D5. Summary of estimation results for Supplemental Study A: Hypothetical ACBC

	$\hat{\boldsymbol{b}}$ and $\hat{b}^{p*}$		$\hat{\boldsymbol{\sigma}}$ and $\hat{\sigma}^{p*}$	
Parameter	**Mean**	**95% CI**	**Mean**	**95% CI**
None	4.662	[3.943, 5.439]	1.961	[1.491, 2.500]
500GB	-0.730	[-1.026, -0.450]	1.109	[0.849, 1.426]
Black	0.262	[0.079, 0.460]	0.627	[0.411, 0.875]
One dual shock controller	-1.407	[-1.786, -1.061]	1.303	[0.990, 1.681]
Charging station	0.735	[0.450, 1.026]	0.960	[0.677, 1.275]
Wireless headset	-0.159	[-0.520, 0.165]	1.135	[0.814, 1.518]
Far Cry Primal	-0.234	[-0.730, 0.234]	1.632	[1.238, 2.099]
GTA V	0.912	[0.437, 1.388]	1.802	[1.394, 2.266]
Life is strange	-0.321	[-0.750, 0.087]	1.415	[1.041, 1.832]
TC's The Division	-0.184	[-0.533, 0.141]	0.874	[0.512, 1.266]
Witcher 3	0.274	[-0.081, 0.612]	1.029	[0.682, 1.406]
Fallout 4	0.248	[-0.124, 0.592]	1.123	[0.787, 1.505]
Final Fantasy x/x2	-0.574	[-1.016, -0.151]	1.369	[0.994, 1.787]
Dark Souls 3	0.087	[-0.241, 0.393]	0.575	[0.063, 1.047]
Just Dance 2016	-1.205	[-1.936, -0.555]	2.416	[1.860, 3.059]
Guitar Hero	-0.251	[-0.827, 0.295]	1.911	[1.415, 2.514]
FIFA 16	1.478	[0.773, 2.171]	2.739	[2.191, 3.381]
Price	0.476	[0.174, 0.756]	1.079	[0.854, 1.356]

Notes: $LMD = -1577.528$, $\min(n_{eff}) = 3009$, $\text{median}(n_{eff}) = 4694$, and $\max(\hat{R}) = 1.002$. The implied average price sensitivity and the corresponding standard deviation in the population at the mean of the results are $\hat{b}^{p} = -2.881$ and $\hat{\sigma}^{p} = 4.277$, respectively.

Table D6. Summary of estimation results for Supplemental Study A: Incentive-aligned ACBC

	$\hat{\boldsymbol{b}}$ and $\hat{b}^{p*}$		$\hat{\boldsymbol{\sigma}}$ and $\hat{\sigma}^{p*}$	
Parameter	**Mean**	**95% CI**	**Mean**	**95% CI**
None	4.888	[4.062, 5.746]	2.412	[1.874, 3.021]
500GB	-0.431	[-0.711, -0.164]	1.044	[0.778, 1.364]
Black	0.729	[0.395, 1.084]	1.367	[1.050, 1.751]
One dual shock controller	-1.102	[-1.440, -0.797]	1.202	[0.923, 1.541]
Charging station	0.745	[0.446, 1.052]	1.004	[0.719, 1.354]
Wireless headset	-0.319	[-0.643, -0.019]	0.856	[0.500, 1.245]
Far Cry Primal	0.346	[-0.029, 0.718]	1.231	[0.897, 1.611]
GTA V	0.771	[0.333, 1.214]	1.624	[1.237, 2.069]
Life is Strange	-0.614	[-1.070, -0.177]	1.531	[1.137, 1.997]
TC's The Division	0.214	[-0.204, 0.615]	1.353	[0.960, 1.803]
Witcher 3	0.217	[-0.231, 0.634]	1.396	[0.994, 1.859]
Fallout 4	0.385	[-0.000, 0.768]	1.215	[0.845, 1.631]
Final Fantasy x/x2	-0.508	[-0.961, -0.056]	1.564	[1.163, 2.007]
Dark Souls 3	0.062	[-0.392, 0.499]	1.477	[1.035, 1.987]
Just Dance 2016	-1.445	[-2.143, -0.815]	2.139	[1.592, 2.801]
Guitar Hero	-0.132	[-0.790, 0.469]	2.197	[1.665, 2.824]
FIFA 16	1.202	[0.526, 1.883]	2.700	[2.153, 3.343]
Price	0.569	[0.272, 0.841]	1.055	[0.838, 1.337]

Notes: $LMD = -1529.604$, $\min(n_{eff}) = 3958$, $\text{median}(n_{eff}) = 4842$, and $\max(\hat{R}) = 1.002$. The implied average price sensitivity and the corresponding standard deviation in the population at the mean of the results are $\hat{b}^p = -3.082$ and $\hat{\sigma}^p = 4.406$, respectively.

Table D7. Summary of estimation results for Study 2: Hypothetical CBC

	$\hat{\boldsymbol{b}}$ and $\hat{b}^{p*}$		$\hat{\boldsymbol{\sigma}}$ and $\hat{\sigma}^{p*}$	
Parameter	**Mean**	**95% CI**	**Mean**	**95% CI**
None	-0.499	[-1.126, 0.088]	2.620	[2.089, 3.307]
White	0.021	[-0.167, 0.196]	0.577	[0.367, 0.808]
Black	0.164	[-0.034, 0.366]	0.722	[0.511, 0.953]
900 Watt	-0.237	[-0.380, -0.101]	0.480	[0.332, 0.649]
Plastic mixing bowl	-0.262	[-0.469, -0.068]	0.426	[0.055, 0.756]
Stainless steel mixing bowl, polished	0.431	[0.220, 0.636]	0.566	[0.304, 0.841]
Stainless steel mixing bowl, brushed	0.646	[0.410, 0.883]	0.774	[0.530, 1.041]
Disc for potatoes	0.257	[0.079, 0.436]	0.293	[0.021, 0.567]
Disc for vegetables	-0.131	[-0.310, 0.048]	0.254	[0.015, 0.566]
Discs for potatoes and vegetables	0.507	[0.332, 0.693]	0.226	[0.012, 0.522]
Measurement cup	0.264	[0.162, 0.370]	0.151	[0.009, 0.334]
Mincer	0.123	[-0.055, 0.299]	0.246	[0.014, 0.550]
Mincer with shortbread biscuits attachment	0.315	[0.142, 0.497]	0.234	[0.016, 0.518]
Mincer with grater	0.100	[-0.078, 0.280]	0.248	[0.013, 0.551]
Citrus juicer	-0.227	[-0.472, 0.019]	0.334	[0.019, 0.738]
Blender attachment	0.231	[0.009, 0.458]	0.236	[0.013, 0.601]
Ice maker	0.307	[0.068, 0.550]	0.461	[0.041, 0.875]
TastyMoments	0.535	[0.241, 0.826]	0.921	[0.591, 1.267]
Blender	0.301	[0.027, 0.563]	0.701	[0.268, 1.092]
Smoothies and Shakes	0.096	[-0.130, 0.326]	0.172	[0.006, 0.467]
Vegetarian	0.089	[-0.151, 0.334]	0.429	[0.044, 0.794]
Low carb	0.090	[-0.139, 0.317]	0.177	[0.007, 0.482]
Sweet and easy	-0.007	[-0.258, 0.234]	0.380	[0.032, 0.768]
Jamie's 5-ingredients	-0.052	[-0.281, 0.173]	0.198	[0.010, 0.532]
Price	-0.395	[-0.842, -0.014]	1.333	[1.010, 1.745]

Notes: LMD $= -1163.437$, $\min(n_{\mathrm{eff}}) = 2711$, $\mathrm{median}(n_{\mathrm{eff}}) = 4815$, and $\max(\hat{R}) = 1.002$. The implied average price sensitivity and the corresponding standard deviation in the population at the mean of the results are $\hat{b}^{p} = -1.638$ and $\hat{\sigma}^{p} = 3.630$, respectively.

Table D8. Summary of estimation results for Study 2: Incentive-aligned CBC

	$\hat{\boldsymbol{b}}$ and $\hat{b}^{p*}$		$\hat{\boldsymbol{\sigma}}$ and $\hat{\sigma}^{p*}$	
Parameter	**Mean**	**95% CI**	**Mean**	**95% CI**
None	-0.224	[-1.028, 0.534]	3.500	[2.781, 4.376]
White	-0.085	[-0.276, 0.103]	0.554	[0.307, 0.808]
Black	0.243	[0.022, 0.462]	0.779	[0.551, 1.034]
900 Watt	-0.285	[-0.448, -0.127]	0.559	[0.384, 0.759]
Plastic mixing bowl	-0.355	[-0.554, -0.161]	0.183	[0.007, 0.476]
Stainless steel mixing bowl, polished	0.570	[0.383, 0.770]	0.193	[0.009, 0.481]
Stainless steel mixing bowl, brushed	0.625	[0.414, 0.849]	0.495	[0.148, 0.802]
Disc for potatoes	0.197	[0.001, 0.393]	0.299	[0.020, 0.630]
Disc for vegetables	-0.029	[-0.232, 0.169]	0.274	[0.015, 0.608]
Discs for potatoes and vegetables	0.524	[0.305, 0.745]	0.519	[0.171, 0.828]
Measurement cup	0.244	[0.106, 0.388]	0.445	[0.254, 0.638]
Mincer	-0.176	[-0.399, 0.051]	0.605	[0.273, 0.905]
Mincer with shortbread biscuits attachment	0.629	[0.413, 0.858]	0.483	[0.097, 0.807]
Mincer with grater	0.278	[0.080, 0.478]	0.341	[0.024, 0.675]
Citrus juicer	-0.481	[-0.834, -0.161]	0.866	[0.377, 1.309]
Blender attachment	0.130	[-0.147, 0.400]	0.512	[0.039, 1.006]
Ice maker	0.282	[-0.068, 0.626]	1.167	[0.783, 1.595]
TastyMoments	0.486	[0.145, 0.829]	1.160	[0.783, 1.568]
Blender	0.624	[0.371, 0.888]	0.378	[0.023, 0.835]
Smoothies and shakes	0.026	[-0.229, 0.279]	0.222	[0.011, 0.589]
Vegetarian	0.213	[-0.029, 0.461]	0.218	[0.009, 0.578]
Low carb	0.154	[-0.086, 0.399]	0.207	[0.008, 0.553]
Sweet and easy	0.034	[-0.223, 0.287]	0.180	[0.007, 0.494]
Jamie's 5-ingredients	-0.018	[-0.288, 0.246]	0.414	[0.036, 0.823]
Price	-0.111	[-0.503, 0.218]	1.154	[0.885, 1.506]

Notes: $LMD = -1069.936$, $\min(n_{eff}) = 2834$, $\text{median}(n_{eff}) = 4782$, and $\max(\hat{R}) = 1.002$. The implied average price sensitivity and the corresponding standard deviation in the population at the mean of the results are $\hat{b}^p = -1.742$ and $\hat{\sigma}^p = 2.908$, respectively.

Table D9. Summary of estimation results for Study 2: Hypothetical ACBC

	$\hat{\boldsymbol{b}}$ and $\hat{b}^{p*}$		$\hat{\boldsymbol{\sigma}}$ and $\hat{\sigma}^{p*}$	
Parameter	**Mean**	**95% CI**	**Mean**	**95% CI**
None	4.237	[3.526, 4.977]	2.483	[2.016, 3.016]
White	0.174	[-0.095, 0.431]	1.097	[0.823, 1.408]
Black	0.436	[0.164, 0.708]	1.072	[0.804, 1.378]
900 Watt	-0.504	[-0.669, -0.352]	0.626	[0.452, 0.820]
Plastic mixing bowl	-0.360	[-0.602, -0.126]	0.759	[0.508, 1.031]
Stainless steel mixing bowl, polished	0.386	[0.165, 0.595]	0.602	[0.313, 0.886]
Stainless steel mixing bowl, brushed	0.688	[0.471, 0.903]	0.666	[0.433, 0.919]
Disc for potatoes	0.072	[-0.124, 0.255]	0.397	[0.066, 0.690]
Disc for vegetables	-0.195	[-0.402, 0.000]	0.379	[0.042, 0.688]
Discs for potatoes and vegetables	0.759	[0.563, 0.961]	0.595	[0.336, 0.855]
Measurement cup	0.050	[-0.062, 0.160]	0.353	[0.123, 0.554]
Mincer	-0.094	[-0.373, 0.154]	0.677	[0.375, 0.996]
Mincer with shortbread biscuits attachment	0.371	[0.039, 0.668]	1.123	[0.819, 1.487]
Mincer with grater	0.054	[-0.288, 0.369]	0.984	[0.679, 1.351]
Citrus juicer	-0.702	[-1.036, -0.389]	0.947	[0.622, 1.318]
Blender attachment	0.367	[0.063, 0.647]	0.985	[0.696, 1.304]
Ice maker	-0.005	[-0.419, 0.373]	1.482	[1.136, 1.879]
TastyMoments	0.307	[-0.053, 0.643]	1.169	[0.844, 1.543]
Blender	0.868	[0.560, 1.165]	0.948	[0.616, 1.302]
Smoothies and shakes	-0.029	[-0.239, 0.181]	0.248	[0.012, 0.573]
Vegetarian	-0.299	[-0.552, -0.067]	0.484	[0.094, 0.821]
Low carb	0.089	[-0.160, 0.323]	0.621	[0.333, 0.905]
Sweet and easy	-0.176	[-0.448, 0.075]	0.536	[0.133, 0.883]
Jamie's 5-ingredients	0.055	[-0.218, 0.320]	0.597	[0.158, 0.975]
Price	0.522	[0.227, 0.809]	1.256	[1.025, 1.542]

Notes: $\text{LMD} = -3170.964$, $\min(\text{n}_{\text{eff}}) = 2119$, $\text{median}(\text{n}_{\text{eff}}) = 4639$, and $\max(\hat{\text{R}}) = 1.003$. The implied average price sensitivity and the corresponding standard deviation in the population at the mean of the results are $\hat{b}^p = -3.709$ and $\hat{\sigma}^p = 7.271$, respectively.

Table D10. Summary of estimation results for Study 2: Incentive-aligned ACBC

	$\hat{\boldsymbol{b}}$ and $\hat{b}^{p*}$		$\hat{\boldsymbol{\sigma}}$ and $\hat{\sigma}^{p*}$	
Parameter	**Mean**	**95% CI**	**Mean**	**95% CI**
None	4.716	[3.979, 5.488]	2.025	[1.525, 2.590]
White	0.024	[-0.188, 0.228]	0.689	[0.428, 0.959]
Black	0.155	[-0.117, 0.428]	1.065	[0.798, 1.366]
900 Watt	-0.535	[-0.776, -0.307]	1.013	[0.779, 1.277]
Plastic mixing bowl	-0.378	[-0.721, -0.058]	1.257	[0.944, 1.603]
Stainless steel mixing bowl, polished	0.775	[0.532, 1.014]	0.697	[0.382, 1.025]
Stainless steel mixing bowl, brushed	0.915	[0.592, 1.241]	1.250	[0.953, 1.594]
Disc for potatoes	0.105	[-0.095, 0.299]	0.341	[0.029, 0.682]
Disc for vegetables	-0.246	[-0.492, -0.022]	0.569	[0.272, 0.852]
Discs for potatoes and vegetables	0.828	[0.575, 1.081]	0.903	[0.627, 1.198]
Measurement cup	0.128	[-0.007, 0.259]	0.453	[0.283, 0.633]
Mincer	-0.001	[-0.283, 0.264]	0.793	[0.469, 1.136]
Mincer with shortbread biscuits attachment	0.561	[0.228, 0.885]	1.252	[0.924, 1.628]
Mincer with grater	0.505	[0.167, 0.832]	1.151	[0.825, 1.533]
Citrus juicer	-0.579	[-0.961, -0.229]	1.052	[0.697, 1.457]
Blender attachment	0.415	[0.064, 0.739]	1.179	[0.832, 1.566]
Ice maker	0.042	[-0.432, 0.504]	1.713	[1.306, 2.188]
TastyMoments	0.563	[0.086, 1.024]	1.844	[1.445, 2.309]
Blender	1.179	[0.781, 1.564]	1.448	[1.104, 1.849]
Smoothies and shakes	-0.204	[-0.477, 0.061]	0.647	[0.346, 0.953]
Vegetarian	-0.267	[-0.541, -0.011]	0.478	[0.071, 0.846]
Low carb	0.245	[-0.011, 0.483]	0.584	[0.223, 0.917]
Sweet and easy	-0.338	[-0.629, -0.063]	0.641	[0.301, 0.979]
Jamie's 5-ingredients	0.152	[-0.119, 0.411]	0.503	[0.092, 0.860]
Price	0.614	[0.354, 0.853]	1.038	[0.836, 1.278]

Notes: $LMD = -2819.608$, $\min(n_{eff}) = 2238$, $\text{median}(n_{eff}) = 4629$, and $\max(\hat{R}) = 1.003$. The implied average price sensitivity and the corresponding standard deviation in the population at the mean of the results are $\hat{b}^p = -3.167$ and $\hat{\sigma}^p = 4.408$, respectively.

Table D11. Summary of estimation results for Study 3: Hypothetical ACBC

	$\hat{\boldsymbol{b}}$ and $\hat{b}^{p*}$		$\hat{\boldsymbol{\sigma}}$ and $\hat{\sigma}^{p*}$	
Parameter	**Mean**	**95% CI**	**Mean**	**95% CI**
None	3.348	[2.874, 3.861]	1.714	[1.362, 2.110]
Fitbit	0.479	[.286, .680]	0.694	[.477, .934]
TomTom	-0.222	[-.370, -.084]	0.263	[.021, .509]
GPS not included	-2.874	[-3.464, -2.323]	2.261	[1.816, 2.795]
Connected GPS	0.531	[.241, .823]	1.084	[.807, 1.401]
Heart-rate monitor not included	-2.593	[-3.156, -2.100]	2.078	[1.632, 2.608]
Heart-rate monitor chest strap	-0.440	[-.832, -.075]	1.695	[1.347, 2.116]
sleep detection not included	-0.944	[-1.278, -.617]	1.523	[1.227, 1.866]
Food tracker app not included	-0.119	[-.259, .022]	0.520	[.372, .686]
Fitness coach app not included	0.056	[-.080, .195]	0.499	[.339, .678]
Price	1.139	[.901, 1.353]	0.771	[.597, .985]

Notes: $LMD = -1700.25$, $\min(n_{eff}) = 894$, $\text{median}(n_{eff}) = 4142$, and $\max(\hat{R}) = 1.01$. The implied average price sensitivity and the corresponding standard deviation in the population at the mean of the results are $\hat{b}^p = -4.205$ and $\hat{\sigma}^p = 3.789$, respectively.

Table D12. Summary of estimation results for Study 3: Incentive-aligned ACBC

	$\hat{\boldsymbol{b}}$ and $\hat{b}^{p*}$		$\hat{\boldsymbol{\sigma}}$ and $\hat{\sigma}^{p*}$	
Parameter	**Mean**	**95% CI**	**Mean**	**95% CI**
None	4.699	[4.074, 5.374]	1.958	[1.501, 2.489]
Fitbit	0.300	[.132, .471]	0.518	[.295, .744]
TomTom	-0.097	[-.245, .045]	0.181	[.009, .409]
GPS not included	-3.665	[-4.446, -2.985]	2.535	[1.966, 3.171]
Connected GPS	0.203	[-.230, .647]	1.685	[1.324, 2.104]
Heart-rate monitor not included	-3.077	[-3.776, -2.436]	2.432	[1.889, 3.077]
Heart-rate monitor chest strap	-0.907	[-1.498, -.348]	2.395	[1.908, 2.986]
Sleep detection not included	-0.861	[-1.221, -.526]	1.589	[1.273, 1.955]
Food tracker app not included	0.238	[.013, .464]	1.036	[.822, 1.282]
Fitness coach app not included	0.164	[-.053, .385]	0.984	[.767, 1.229]
Price	0.983	[.717, 1.217]	0.800	[.592, 1.049]

Notes: $LMD = -1494.35$, $\min(n_{eff}) = 955$, $\text{median}(n_{eff}) = 4467$, and $\max(\hat{R}) = 1.007$. The implied average price sensitivity and the corresponding standard deviation in the population at the mean of the results are $\hat{b}^p = -3.680$ and $\hat{\sigma}^p = 3.485$, respectively.

Partworth estimates and relative importances

Tables D13 to D16 present the mean partworth estimates of the lower model and corresponding standard deviations of each study's conditions side-by-side for better comparability of results. In general, the mean values of the partworths are very similar to the population means reported in Tables D1 to D12 as all our categorical attributes are effect-coded. For the partworths, we also report the corresponding values for the reference levels (e.g., Pizza size of 29 cm in Study 1), the transformed price parameters to aid the interpretability of the results, and the results for the none parameter after four stages. Note that the standard deviations of the partworths are typically slightly smaller than the estimated population values, as the former only account for the estimated heterogeneity in the sample (Pachali et al. 2020).

Additionally, the tables contain the estimated relative importance of each attribute, which is usually reported in the conjoint literature (Rao 2014). For the relative importance, we calculate for each participant and attribute a the range of the partworths $\hat{\beta}_{i,a}$ across levels m and rescale them such that they add up to 100%:

$$\text{relative importance}_{i,a} = \frac{\max_{l \in M_a}\left(\hat{\beta}_{i,a}\right) - \min_{l \in M_a}\left(\hat{\beta}_{i,a}\right)}{\sum_{a' \in A}\left(\max_{l \in M_{a'}}\left(\hat{\beta}_{i,a'}\right) - \min_{l \in M_{a'}}\left(\hat{\beta}_{i,a'}\right)\right)}. \qquad \text{(formula D9)}$$

Tables D1 to D12 report the average values across participants for each attribute. The relative importance is useful in comparing different conjoint methods, because it does not rely on the scale of partworth utilities. Still, these may obfuscate consumers' preference heterogeneity within conditions. It should be noted that comparing the relative importance values between CBC and ACBC should be taken cautiously, due to the ACBC employing a wider price range than the CBC. This difference in price ranges is a result of how the price attribute was specified; while the CBC's price attribute was based on predefined price levels randomly assigned to each product concept, the ACBC's price attribute was composed of a predefined base price plus the prices of all attribute levels of a product concept (i.e., summed pricing), with random price variation lying within the specified range of -30% to +30%. This has led to a larger range of prices in ACBCs, which partially contributed to larger estimated relative importance as compared to the CBCs. The rank order of attribute importance and partworth within an attribute might, therefore, provide a better basis for comparison. Although there are some changes in the rank orders, the comparison of the rank order of attribute importance and partworths did not reveal any systematic differences between our tested conditions that would lead to a conclusive interpretation.

By tendency, it can be argued that for Study 1 (pizza) the application of adaptive designs (vs. static CBC designs) leads to a less extreme estimated preference for unusual recipe components on a pizza, such as garlic or onion rings. Some parallels can be spotted in Supplemental Study A (PlayStation4 bundles). Again, here the application of adaptive (vs. static) designs in conjoint leads to a lower preference for the non-standard color white (instead of the standard black). By interpretation, one could also argue that the attribute levels that do not belong to the core product usage involved and that are rather auxiliary are estimated to be less liked by participants when applying adaptive instead of static CBC designs. For example, in Supplemental Study A, having a wireless headset with the PlayStation 4 bundle is less liked in the ACBC conditions as compared to CBC. Similarly, in Study 2 (food processors), it seems to be worth more for participants in the ACBC conditions (vs. CBC) to have none of the offered additional recipe books to be included in the food processor bundle. Study 3 (fitness trackers) was the only study with competition among varying brands (Fitbit vs. TomTom vs. Withings). It could be noted that with the application of incentive-aligned ACBC (vs. hypothetical ACBC) brand competition is losing importance (least important attribute in incentive-aligned ACBC vs. fifth most important attribute in hypothetical ACBC).

Table D13. Partworth estimates and relative importance: Study 1

	Incentive-aligned CBC		Hypothetical ACBC		Incentive-aligned ACBC	
	Mean	***SD***	**Mean**	***SD***	**Mean**	***SD***
None (3 stages)	3.419	3.292	3.784	1.462	4.602	2.423
None (4 stages)			5.015	2.581	4.586	3.188
Pizza type	26.2%		19.1%		19.9%	
Margherita	-0.918	1.265	-0.671	2.010	-0.234	1.671
Salami	0.495	1.076	-0.052	2.404	0.315	1.475
Funghi	-0.229	1.536	-0.726	2.356	-1.330	2.406
Salami, ham & funghi	-0.145	1.382	0.669	1.978	0.096	2.109
Chicken	0.360	1.309	-0.235	1.862	0.640	2.039
Four cheeses	0.178	1.432	0.892	1.512	0.612	1.062
Tomato	0.616	1.196	0.733	1.847	0.730	1.602
Gyross	-0.356	2.686	-0.610	2.295	-0.829	2.510
Pizza size	6.6%		8.0%		6.3%	
20cm	-0.604	0.399	-1.426	1.037	-0.913	0.888
25cm	0.047	0.079	0.479	0.166	0.420	0.140
29cm	0.556	0.409	0.947	1.029	0.492	0.857
Additional sauces	8.8%		10.8%		9.3%	
Hollandaise	-0.458	0.792	-0.894	1.303	-0.712	1.189
BBQ	-0.064	0.501	-0.720	0.962	-0.614	0.999
None	0.522	0.838	1.615	1.162	1.326	1.037
Additional toppings	16.0%		14.3%		13.5%	
Onion rings	0.545	0.374	0.163	1.224	-0.177	1.204
Paprika	0.383	0.240	0.682	0.857	0.848	0.316
Jalapenos	0.372	0.681	0.337	1.418	0.284	1.237
Olives	-0.353	0.900	-0.781	1.622	-0.870	1.296
Boiled egg	-1.422	0.932	-1.495	1.874	-1.469	1.695
None	0.475	0.960	1.095	1.011	1.385	0.970
Additional cheese	6.0%		3.2%		3.0%	
Gouda	0.227	0.073	-0.308	0.186	-0.026	0.276
Feta	-0.086	0.389	-0.025	0.496	-0.446	0.184
Mozzarella	0.290	0.307	0.334	0.192	0.368	0.075
None	-0.431	0.568	-0.001	0.474	0.103	0.396
Spices	3.2%		1.5%		1.7%	
Garlic	0.260	0.202	-0.061	0.292	-0.077	0.319
None	-0.260	0.202	0.061	0.292	0.077	0.319
Beverages	6.7%		7.9%		4.8%	
Coca-Cola	0.428	0.309	0.514	0.725	0.011	0.683
Coke light or zero	0.469	0.147	-0.447	1.177	-0.089	0.411
Fanta	-0.236	0.096	0.052	0.306	0.120	0.205
Sprite	-0.140	0.174	0.049	0.474	-0.192	0.202
Water (still or medium)	-0.112	0.133	0.198	0.832	0.077	0.502
None	-0.409	0.547	-0.367	1.014	0.075	0.946
Price (importance)	26.5%		35.2%		41.5%	
Price (utility)	-6.287	3.935	-6.740	3.509	-8.741	5.398

Notes: Grey shaded rows show the attributes and their relative importance and white rows the means and the standard deviations of the partworths. The price range (maximum price – minimum price) was €8.00 in the incentive-aligned CBC, €17.85 in the hypothetical ACBC, and €17.00 in the incentive-aligned ACBC.

Table D14. Partworth estimates and relative importance: Supplemental Study A

	Incentive-aligned CBC		**Hypothetical ACBC**		**Incentive-aligned ACBC**	
	Mean	***SD***	**Mean**	***SD***	**Mean**	***SD***
None (3 stages)	3.030	1.915	4.671	1.574	4.905	1.975
None (4 stages)			6.223	3.711	6.725	5.845
Hard disc capacity	5.5%		6.2%		4.8%	
500GB	-0.448	0.311	-0.729	0.922	-0.429	0.829
1TB	0.448	0.311	0.729	0.922	0.429	0.829
Color	2.9%		2.8%		6.5%	
Black	-0.041	0.317	0.261	0.458	0.734	1.154
White	0.041	0.317	-0.261	0.458	-0.734	1.154
Number of controllers	6.2%		9.7%		7.3%	
One dual shock controller	-0.430	0.562	-1.409	1.077	-1.103	1.005
Two dual shock controllers	0.430	0.562	1.409	1.077	1.103	1.005
Accessories	3.9%		7.2%		6.2%	
Charging station	0.160	0.047	0.736	0.704	0.743	0.750
Wireless headset	0.142	0.278	-0.161	0.867	-0.320	0.572
None	-0.302	0.296	-0.575	1.048	-0.422	0.868
Action-adventure games	15.8%		11.3%		11.1%	
Far Cry Primal	0.391	0.401	-0.238	1.307	0.347	0.974
GTA V	0.878	0.915	0.917	1.556	0.769	1.374
Life is Strange	0.040	0.659	-0.327	1.158	-0.616	1.153
T.C.'s The Division	-0.116	0.812	-0.182	0.554	0.216	1.024
None	-1.193	0.866	-0.169	1.066	-0.716	1.026
Role-playing games	12.1%		8.9%		10.4%	
Witcher 3	0.351	0.852	0.274	0.694	0.220	1.074
Fallout 4	0.538	0.254	0.246	0.799	0.390	0.896
Final Fantasy X/X2	-0.329	0.638	-0.575	1.055	-0.514	1.247
Dark Souls 3	0.174	0.456	0.085	0.273	0.061	1.122
None	-0.734	0.953	-0.030	1.353	-0.157	1.537
Games for family and companionship	16.5%		18.0%		16.8%	
Just Dance 2016	-0.497	1.179	-1.212	2.135	-1.455	1.687
Guitar Hero	-0.050	1.040	-0.253	1.529	-0.137	1.831
FIFA 16	0.724	1.317	1.482	2.602	1.208	2.470
None	-0.176	1.060	-0.017	1.835	0.383	1.893
Price (importance)	37.1%		36.1%		36.8%	
Price (utility)	-1.684	0.802	-2.519	2.059	-2.732	2.354

Notes: Grey shaded rows rows show the attributes and their relative importance and white rows the means and the standard deviations of the partworths. The price range (maximum price – minimum price) was €400.00 in the incentive-aligned CBC, €528.00 in the hypothetical ACBC, and €510.00 in the incentive-aligned ACBC.

Table D15. Partworth estimates and relative importance: Study 2

	Hypothetical CBC		**Incentive-aligned CBC**		**Hypothetical ACBC**		**Incentive-aligned ACBC**	
	Mean	***SD***	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***
None (3 stages)	-0.499	2.252	-0.223	3.101	4.238	2.160	4.723	1.567
None (4 stages)					4.408	4.881	5.274	3.587
Color	9.4%		9.0%		9.6%		6.5%	
White	0.019	0.365	-0.086	0.333	0.174	0.840	0.023	0.469
Black	0.165	0.499	0.243	0.545	0.436	0.816	0.154	0.840
Red	-0.184	0.592	-0.157	0.712	-0.610	1.162	-0.177	1.036
Power	6.2%		6.0%		5.2%		5.7%	
900 Watt	-0.236	0.311	-0.287	0.377	-0.505	0.464	-0.537	0.850
1000 Watt	0.236	0.311	0.287	0.377	0.505	0.464	0.537	0.850
Additional mixing bowl	15.8%		12.7%		8.3%		11.5%	
Plastic	-0.261	0.205	-0.356	0.049	-0.361	0.516	-0.381	0.983
Stainless steel polished	0.430	0.342	0.570	0.053	0.385	0.355	0.774	0.447
Stainless steel brushed	0.647	0.535	0.624	0.252	0.689	0.442	0.917	1.002
None	-0.816	0.664	-0.838	0.281	-0.713	0.589	-1.311	1.023
Additional discs	11.3%		10.4%		7.1%		7.3%	
Disc for potatoes	0.256	0.123	0.197	0.113	0.072	0.189	0.105	0.152
Disc for vegetables	-0.131	0.093	-0.029	0.093	-0.195	0.177	-0.246	0.321
Both	0.506	0.080	0.524	0.275	0.759	0.388	0.830	0.669
None	-0.632	0.217	-0.692	0.370	-0.636	0.406	-0.689	0.756
Measuring cup	5.3%		4.6%		1.6%		2.1%	
Included	0.264	0.053	0.243	0.265	0.050	0.204	0.127	0.303
Not included	-0.264	0.053	-0.243	0.265	-0.050	0.204	-0.127	0.303
Mincer	8.3%		12.5%		10.5%		11.2%	
Included	0.123	0.087	-0.178	0.341	-0.093	0.450	0.000	0.501
Included with a shortbread biscuit attachment	0.315	0.090	0.629	0.239	0.370	0.869	0.563	0.956
Included with a grater	0.099	0.087	0.278	0.135	0.055	0.685	0.504	0.831
Not included	-0.536	0.199	-0.730	0.480	-0.331	1.569	-1.067	1.626
Further attachments	18.4%		19.2%		13.1%		15.6%	
Citrus juicer	-0.227	0.124	-0.480	0.474	-0.704	0.569	-0.579	0.686
Blender attachment	0.231	0.068	0.131	0.216	0.366	0.671	0.413	0.843
Ice maker	0.305	0.206	0.281	0.770	-0.009	1.106	0.040	1.334
TastyMoments	0.534	0.600	0.485	0.765	0.310	0.849	0.562	1.466
Blender	0.300	0.375	0.624	0.143	0.869	0.614	1.181	1.123
None	-1.143	0.715	-1.041	0.791	-0.832	0.952	-1.617	1.378
Recipe book	5.5%		5.8%		5.0%		5.1%	
Smoothies and shakes	0.096	0.036	0.027	0.056	-0.028	0.085	-0.205	0.382
Vegetarian	0.090	0.192	0.212	0.057	-0.301	0.240	-0.266	0.228
Low carb	0.091	0.040	0.153	0.053	0.090	0.353	0.244	0.346
Sweet and easy	-0.007	0.158	0.034	0.039	-0.177	0.272	-0.336	0.371
Jamie's 5-ingredients	-0.052	0.049	-0.017	0.176	0.054	0.309	0.153	0.243
None	-0.218	0.333	-0.410	0.221	0.361	0.514	0.411	0.624
Price (importance)	19.8%		19.9%		39.7%		34.9%	
Price (utility)	-1.402	1.609	-1.646	2.105	-3.011	2.744	-2.847	2.404

Notes: Grey shaded rows show the attributes and their relative importance and white rows the means and the standard deviations of the partworths. The price range (maximum price – minimum price) was €175.00 in the hypothetical and the incentive-aligned CBC, €386.00 in the hypothetical ACBC, and €377.00 in the incentive-aligned ACBC.

Table D16. Partworth estimates and relative importance: Study 3

	Hypothetical ACBC		Incentive-aligned ACBC	
	Mean	***SD***	**Mean**	***SD***
None (3 stages)	3.355	1.404	4.705	1.532
None (4 stages)	6.317	4.775	8.071	4.142
Brand	4.5%		2.3%	
Fitbit	0.479	0.510	0.300	0.338
TomTom	-0.222	0.131	-0.097	0.081
Withings	-0.257	0.482	-0.203	0.334
GPS	20.6%		24.4%	
No GPS	-2.885	1.924	-3.677	2.099
Connected GPS	0.535	0.817	0.201	1.378
Integrated GPS	2.350	1.868	3.476	1.881
Heart rate monitor (HRM)	22.0%		25.7%	
No HRM	-2.600	1.712	-3.087	2.008
HRM – chest strap	-0.441	1.418	-0.900	2.038
Integrated HRM	3.041	1.973	3.987	2.163
Sleep detection	9.5%		8.3%	
Without sleep detection	-0.947	1.334	-0.861	1.414
With sleep detection	0.947	1.334	0.861	1.414
Food tracker app	2.7%		4.6%	
Not included	-0.118	0.393	0.237	0.892
Included	0.118	0.393	-0.237	0.892
Fitness coach app	2.3%		4.3%	
Not included	0.056	0.365	0.165	0.837
Included	-0.056	0.365	-0.165	0.837
Price (importance)	38.4%		30.4%	
Price (utility)	-4.115	2.657	-3.585	2.645

Notes: Grey shaded rows show rows show the attributes and their relative importance and white rows the means and the standard deviations of the partworths. The price range (maximum price – minimum price) was €270.00 in the hypothetical ACBC and €272.00 in the incentive-aligned ACBC.

Willingness-to-pay

To make the partworth utilities more interpretable (i.e., "scale-free"), we calculated the willingness-to-pay (WTP)[4] for each attribute level as described in Allenby et al. (2014). For this calculation, we used the individual-level estimates from the lower model and divided the partworths by the absolute value of the price parameter (Train 2009):

$$\mathrm{wtp}_{i,a} = \frac{\hat{\beta}_{i,a}}{\left|\hat{\beta}_i^p\right|} \qquad \text{(formula D10)}$$

Tables D17 to D20 show the estimated median WTP (and median absolute deviation; MAD) for each attribute level (Sonnier et al. 2007). The difference between the monetary values of two levels within each attribute can be interpreted as the price premium participants were implicitly willing to pay for an upgrade from a reference level to another level if the difference is positive or as the price discount that they were implicitly demanding for a downgrade from one level to another if the difference is negative.

In general, we find that most WTP values are reasonable (note the different scaling of the price attribute in the different studies). For example, participants are willing to pay about €0.4 to €2.1 more for a salami pizza over a Margherita, about €37 to €51 for a popular action-adventure game like GTA V (vs. no game in this category for the bundle), €30 to €40 more for a food processor with more power (900 vs. 1000 Watt). The price premium for a fitness tracker from FitBit vs. Withings is about €14 to €20.

Furthermore, the differences across conditions mentioned before based on the interpretation of partworths are still apparent from the WTP results. For example, for additional garlic, participants are willing to pay about €0.8 in the incentive-aligned CBC condition. In the condition with adaptive designs, the WTP for this attribute is much lower, and the direction is

[4] Later in Web Appendix H, we also analyze reservation prices, which are related to the WTP results. However, instead of only looking at the monetary equivalent of particular attributes and levels, the reservation price is the monetary value of a whole product over not buying at all.

flipped. The median WTP is -€0.2; thus, more than 50% of the participant are willing to pay for not having additional garlic on their pizzas. However, there is a considerable amount of heterogeneity, hence some participants still prefer pizza with garlic. We also observed a lower WTP range in ACBCs than CBCs for attributes that do not belong to the core product, such as *beverages* in Study 1 or *recipe book* in Study 2.

We conclude that the WTP results generally have face validity, but we also find interesting differences across conditions. Given that the incentive-aligned ACBC studies lead to the best predictions, we consider the WTP results from this condition as the most relevant for managerial decision-making.

Table D17. Summary of estimated willingness-to-pay: Study 1

WTP in €10	Incentive-aligned CBC		Hypothetical ACBC		Incentive-aligned ACBC	
	Median	***MAD***	**Median**	***MAD***	**Median**	***MAD***
None (3 stages)	0.554	0.419	0.652	0.280	0.569	0.227
None (4 stages)			0.853	0.340	0.621	0.300
Pizza type						
Margherita	-0.134	0.105	-0.109	0.226	-0.018	0.136
Salami	0.072	0.115	-0.024	0.259	0.025	0.124
Funghi	-0.035	0.173	-0.097	0.260	-0.139	0.260
Salami, ham & funghi	-0.061	0.160	0.136	0.187	-0.023	0.212
Chicken	0.032	0.174	-0.041	0.208	0.040	0.167
Four cheeses	0.021	0.130	0.112	0.168	0.061	0.100
Tomato	0.073	0.129	0.120	0.224	0.073	0.148
Gyros	-0.157	0.371	-0.050	0.333	-0.102	0.274
Pizza size						
20cm	-0.107	0.064	-0.233	0.125	-0.121	0.105
25cm	0.007	0.006	0.082	0.030	0.051	0.026
29cm	0.096	0.061	0.118	0.116	0.064	0.078
Additional sauces						
Hollandaise	-0.063	0.094	-0.134	0.176	-0.111	0.106
BBQ	-0.022	0.044	-0.110	0.125	-0.064	0.080
None	0.091	0.095	0.224	0.174	0.150	0.091
Additional toppings						
Onion rings	0.100	0.059	0.003	0.117	-0.011	0.101
Paprika	0.068	0.039	0.116	0.110	0.099	0.039
Jalapenos	0.059	0.078	0.036	0.157	0.020	0.119
Olives	-0.062	0.098	-0.126	0.207	-0.098	0.113
Boiled egg	-0.284	0.195	-0.190	0.274	-0.181	0.177
None	0.067	0.102	0.159	0.137	0.149	0.100
Additional cheese						
Gouda	0.043	0.015	-0.045	0.025	-0.010	0.019
Feta	-0.017	0.037	-0.014	0.058	-0.055	0.029
Mozzarella	0.048	0.043	0.053	0.026	0.048	0.021
None	-0.073	0.077	0.010	0.058	0.011	0.037
Spices						
Garlic	0.042	0.032	-0.011	0.029	-0.010	0.026
None	-0.042	0.032	0.011	0.029	0.010	0.026
Beverages						
Coca-Cola	0.081	0.053	0.068	0.079	-0.001	0.057
Coke light or zero	0.085	0.035	-0.048	0.127	-0.014	0.031
Fanta	-0.047	0.019	0.006	0.027	0.011	0.016
Sprite	-0.021	0.024	0.004	0.047	-0.025	0.017
Water (still or medium)	-0.017	0.018	0.014	0.097	0.000	0.042
None	-0.077	0.080	-0.057	0.115	0.004	0.077

Table D18. Summary of estimated willingness-to-pay: Supplemental Study A

WTP in €100	**Incentive-aligned CBC**		**Hypothetical ACBC**		**Incentive-aligned ACBC**	
	Median	***MAD***	**Median**	***MAD***	**Median**	***MAD***
None (3 stages)	1.939	0.900	2.592	1.448	2.430	1.494
None (4 stages)			2.993	1.543	2.828	1.842
Hard disc capacity						
500GB	-0.287	0.157	-0.211	0.288	-0.100	0.261
1TB	0.287	0.157	0.211	0.288	0.100	0.261
Color						
Black	-0.007	0.119	0.100	0.159	0.197	0.303
White	0.007	0.119	-0.100	0.159	-0.197	0.303
Number of controllers						
One dual shock controller	-0.225	0.232	-0.509	0.440	-0.348	0.320
Two dual shock controllers	0.225	0.232	0.509	0.440	0.348	0.320
Accessories						
Charging station	0.095	0.034	0.276	0.315	0.256	0.250
Wireless headset	0.096	0.114	-0.071	0.261	-0.105	0.159
None	-0.195	0.124	-0.152	0.346	-0.189	0.301
Action-adventure games						
Far Cry Primal	0.240	0.148	-0.123	0.428	0.110	0.321
GTA V	0.501	0.442	0.304	0.415	0.185	0.464
Life is Strange	-0.020	0.255	-0.068	0.383	-0.201	0.360
T.C.'s The Division	-0.064	0.299	-0.091	0.162	0.074	0.360
None	0.130	0.362	-0.111	0.295	-0.325	0.424
Role-playing games						
Witcher 3	0.130	0.362	0.048	0.180	-0.008	0.295
Fallout 4	0.314	0.135	0.040	0.210	0.116	0.300
Final Fantasy X/X2	-0.262	0.264	-0.199	0.382	-0.301	0.354
Dark Souls 3	0.100	0.184	0.014	0.073	-0.014	0.268
None	-0.378	0.362	0.000	0.421	-0.106	0.464
Games for family and companionship						
Just Dance 2016	-0.365	0.294	-0.488	0.753	-0.615	0.718
Guitar Hero	-0.052	0.376	-0.076	0.493	-0.179	0.511
FIFA 16	0.370	0.492	0.568	0.950	0.399	0.755
None	-0.148	0.450	0.013	0.513	0.039	0.557

Table D19. Summary of estimated willingness-to-pay: Study 2

WTP in €100	Hypothetical CBC		Incentive-aligned CBC		Hypothetical ACBC		Incentive-aligned ACBC	
	Median	*MAD*	Median	*MAD*	Median	*MAD*	Median	*MAD*
None (3 stages)	-0.411	1.849	-0.278	1.899	1.751	1.060	2.040	1.113
None (4 stages)					1.783	1.248	1.887	1.084
Color								
White	0.000	0.213	-0.040	0.167	0.014	0.249	0.002	0.153
Black	0.088	0.280	0.144	0.195	0.082	0.252	0.038	0.233
Red	-0.073	0.379	-0.089	0.304	-0.204	0.370	-0.087	0.264
Power								
900 Watt	-0.212	0.188	-0.156	0.201	-0.178	0.184	-0.196	0.221
1000 Watt	0.212	0.188	0.156	0.201	0.178	0.184	0.196	0.221
Additional mixing bowl								
Plastic	-0.249	0.164	-0.324	0.186	-0.139	0.202	-0.151	0.341
Stainless steel polished	0.384	0.329	0.531	0.270	0.148	0.127	0.378	0.297
Stainless steel brushed	0.680	0.557	0.512	0.241	0.202	0.158	0.370	0.336
None	-0.744	0.630	-0.703	0.286	-0.316	0.282	-0.626	0.521
Additional discs								
Disc for potatoes	0.269	0.161	0.179	0.117	0.018	0.048	0.037	0.047
Disc for vegetables	-0.144	0.118	-0.027	0.047	-0.073	0.082	-0.127	0.115
Both	0.628	0.363	0.490	0.263	0.305	0.243	0.397	0.299
None	-0.717	0.394	-0.619	0.375	-0.268	0.227	-0.372	0.363
Measuring cup								
Included	0.333	0.196	0.161	0.137	0.012	0.055	0.033	0.080
Not included	-0.333	0.196	-0.161	0.137	-0.012	0.055	-0.033	0.080
Mincer								
Included	0.112	0.093	-0.102	0.181	-0.016	0.128	-0.012	0.159
Included w. shortbread biscuit attachment	0.369	0.235	0.529	0.316	0.116	0.290	0.224	0.415
Included with a grater	0.070	0.070	0.223	0.128	0.032	0.200	0.169	0.234
Not included	-0.590	0.363	-0.597	0.361	-0.195	0.501	-0.479	0.563
Further attachments								
Citrus juicer	-0.246	0.176	-0.386	0.365	-0.277	0.243	-0.319	0.286
Blender attachment	0.273	0.173	0.095	0.127	0.092	0.188	0.158	0.258
Ice maker	0.277	0.222	0.129	0.404	-0.079	0.256	-0.052	0.368
Tastymoments	0.413	0.425	0.326	0.419	0.069	0.205	0.096	0.427
Blender	0.242	0.241	0.572	0.267	0.312	0.218	0.402	0.361
None	-1.244	0.827	-0.793	0.479	-0.317	0.320	-0.736	0.597
Recipe book								
Smoothies and shakes	0.106	0.068	0.018	0.030	-0.010	0.024	-0.085	0.108
Vegetarian	0.091	0.144	0.202	0.093	-0.127	0.108	-0.107	0.099
Low carb	0.093	0.073	0.132	0.067	0.025	0.081	0.094	0.100
Sweet and easy	-0.019	0.099	0.032	0.031	-0.065	0.096	-0.114	0.137
Jamie’s 5-ingredients	-0.046	0.035	-0.020	0.077	0.006	0.087	0.041	0.065
None	-0.214	0.242	-0.342	0.217	0.108	0.146	0.137	0.188

Table D20. Summary of estimated willingness-to-pay: Study 3

WTP in €100	Hypothetical ACBC		Incentive-aligned ACBC	
	Median	***MAD***	**Median**	***MAD***
None (3 stages)	0.912	0.347	1.570	0.816
None (4 stages)	1.608	0.823	2.666	1.606
Brand				
Fitbit	0.126	0.111	0.080	0.088
TomTom	-0.061	0.040	-0.029	0.025
Withings	-0.075	0.094	-0.058	0.075
GPS				
No GPS	-0.842	0.493	-1.253	0.846
Connected GPS	0.115	0.163	-0.024	0.319
Integrated GPS	0.718	0.555	1.253	0.852
Heart rate monitor (HRM)				
No HRM	-0.670	0.438	-0.873	0.707
HRM – chest strap	-0.109	0.255	-0.281	0.495
Integrated HRM	0.795	0.469	1.209	0.831
Sleep detection				
Without sleep detection	-0.155	0.168	-0.139	0.275
With sleep detection	0.155	0.168	0.139	0.275
Food tracker app				
Not included	-0.028	0.064	0.044	0.190
Included	0.028	0.064	-0.044	0.190
Fitness coach app				
Not included	0.011	0.063	0.065	0.142
Included	-0.011	0.063	-0.065	0.142

Evaluation of predictive validity in terms of hit rate for all main studies

This section summarizes additional model-based results regarding the predictive validity of Studies 1, 2, and 3. Specifically, we present the findings from the analyses where the participant-level hits (and misses) for each study and HOT are used as the dependent variables instead of the MHP, as discussed in the main manuscript. These results are consistent with those reported in the manuscript and further support our conclusions.

Study 1: Pizza menus

We tested for differences in hit rates by applying logistic regression on hit predictions (correct = 1, incorrect = 0) with the condition as the independent variable (reference level = incentive-aligned CBC). Using estimated marginal means (Searle et al. 1980), we additionally assessed the difference between incentive-aligned and hypothetical ACBC. First, the results demonstrate that incentive-aligned ACBC (48.89%) dominates both its hypothetical counterpart (33.33%; $\beta = 0.65$, $z = 2.13$, $p_{(\text{one-tailed})} = 0.017$) and incentive-aligned CBC (22.11%; $\beta = 1.22$, $z = 3.74$, $p_{(\text{one-tailed})} < 0.001$).[5] At the same time, hypothetical ACBC and incentive-aligned CBC perform comparatively well ($\beta = 0.57$, $z = 1.71$, $p = 0.087$).

Study 2: Food processors

As outlined in the manuscript, we applied a generalized, linear mixed-effects model with a random intercept for both, the individuals and the HOTs 1-3 to evaluate the hit rates (binary dependent variable: correct = 1, incorrect = 0, logistic model). The independent variables were, first, incentive-aligned (1) against hypothetical (0) conjoint versions, and second, ACBC (1) against CBC (0). The model elicits a significant positive main effect of incentive alignment ($\beta = 0.41$, $z = 3.37$, $p_{(\text{one-tailed})} < 0.001$) and a positive main effect of adaptive designs ($\beta = 0.64$, $z = 5.25$, $p_{(\text{one-tailed})} < 0.001$) on the hit rates. As for the MHP, additional analysis rejects an interaction of both factors ($\beta = -0.17$, $z = -0.68$, $p = 0.496$). Finally, we find a positive but insignificant difference between hypothetical ACBC and incentive-aligned CBC using estimated marginal means ($\beta = 0.23$, $z = 1.40$, $p = 0.162$).

Study 3: Fitness trackers

Applying logistic regressions predicting the hits (correct = 1, incorrect = 0, logistic model), we found that incentive alignment had a marginally significant effect ($\beta = 0.37$, $z = 1.32$, $p_{(\text{one-tailed})} = 0.093$).

[5] Please note that we report the parameter estimates on the logit-scale across all studies.

Appendix E: Using Stan for estimating HB MNL models: A case study

We present a tutorial on using our code in R and Stan (ver. 2.25, Stan Development Team 2020b) for estimating and analyzing HB MNL models in ACBC studies. We aim to provide a clear explanation and improve accessibility for researchers and applied analysts regarding the functionality of our R/Stan scripts and the relevant data preparation steps (i.e., opening the "black box"). Additionally, we seek to benchmark our Stan implementation of the HB MNL model against Sawtooth Software Lighthouse Studio ver. 9.8.1 (hereafter Sawtooth, Sawtooth Software Inc. 2023). This comparison will help establish confidence in our software by verifying that our data transformations are correct and that the results are (near) identical to Sawtooth. This is an insightful comparison as Sawtooth and Stan use different algorithms ("regular" MCMC vs. HMC) and different priors. To achieve these objectives, we present a small case study that demonstrates the structure and workings of our scripts and allows for direct comparison with Sawtooth. For this case study, we used the data of condition III (incentive-aligned ACBC) from our article's Supplemental Study A (PlayStation 4 bundles). For computations outside Sawtooth, we used R for all data preparation steps, the model estimation, the comparison of results, and the evaluation of the predictive validity. Each of these steps has a separate section below. The complete code and fully reproducible results (https://osf.io/e5v4c/) are available in the replication files as an RMarkdown document.

Overview of the R/Stan code

Before we start with the case study and specific details, we briefly explain the structure of our scripts. Fig. E1 provides a comprehensive overview of the relevant components, including input data and transformed output, which are passed between R functions. This flow chart serves two key purposes for the user: (1) it presents the complete code base, and (2) it explicates the connections between the main building blocks. This helps apply our software to new (A)CBC analyses. Note that we only discuss R functions that are relevant for the user (i.e., functions exposed to the user, but not custom functions called within these functions) and programming steps that are not basic (i.e., we expect the user to be able to set a path to the specific folder and know R fundamentals).

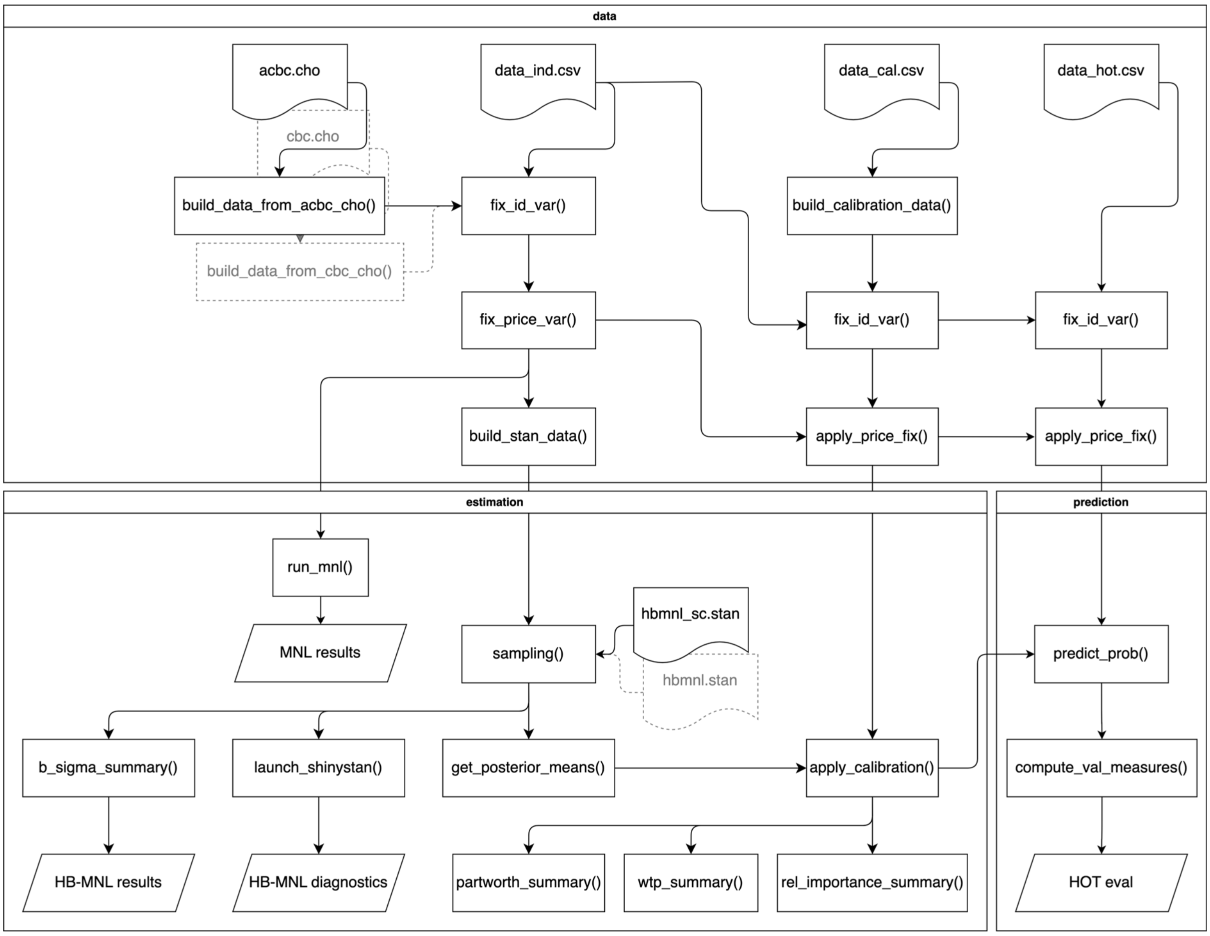


Fig. E1. Flow chart

The R functions and analysis steps can be grouped into three parts: data, estimation, and prediction. The first data part mainly involves transforming data in a general rectangular format (tables in "long" format). While this case study is about ACBC studies, we wrote all R functions and scripts so that CBC data can also be analyzed. The only difference, indicated with dashed lines in Fig. E1, is the first data transformation step (`build_data_from_cbc_cho()` instead of `build_data_from_acbc_cho()`). The estimation part consists of several functions that further transform the output of the main function `sampling()`, which uses the Stan program from the `hbmnl_sc.stan`-file (see Listing D1) and performs the HMC estimation explained in Web Appendix D. In case the user wants to estimate the HB MNL model without sign constraints, as we do in this case study to facilitate comparability with Sawtooth, the Stan file `hbmnl.stan` is relevant (again indicated with dashed lines in Fig. E1). The prediction part is optional (`predict_prob()`) but of specific interest for our research on predictive validity using multiple measures (`compute_val_measures()`) for different conjoint methods.

Data

The data preparation involves three steps: (1) Reshaping Sawtooth's raw data (provided as *.cho-file) into a typical rectangular choice dataset in a "long format." (2) Rescaling of raw prices. (3) Building an additional dataset for the optional fourth stage in an ACBC study (the none parameter calibration step).

The first step is tedious but straightforward and can be automated because all necessary information is included in the *.cho-file. For this purpose, we wrote an R-function (`build_data_from_acbc_cho()`) that manages this step. After reading the *.cho-file into R, the data in the corresponding object looks like Listing E1 (first and last five rows).

Listing E1. Content and structure of *.cho-file

```
        V1 V2 V3 V4 V5 V6 V7     V8 V9
    1:  1  0  9 51  0 NA NA     NA NA
    2:  2  1 NA NA NA NA NA     NA NA
    3:  1  0  0  0  0  0  0 451.26  0
    4:  2  0  0  0  0  0  0 491.26  0
    5:  2 99 NA NA NA NA NA     NA NA
   ---
33371:  3  1 NA NA NA NA NA     NA NA
33372:  1  2  1  1  5  3  4 267.95  0
33373:  1  1  2  3  3  3  2 304.95  0
33374:  2  1  1  3  5  3  2 366.95  0
33375:  2 99 NA NA NA NA NA     NA NA
```

The file has 33,375 rows and nine columns, and the empty cells are pre-filled with NAs. For details, we refer the reader to the Sawtooth manual (Sawtooth Software 2020, chap. 7, App. C). The file contains all relevant information, including participant IDs, as well as details on product attributes and levels from each choice task for the first three ACBC stages. Listing E2 shows the data after the reshaping step.

Listing E2. Content and structure of reshaped choice data

```
       row  obs id cs alt ch none att_11 att_21 ... att_71 att_72 att_73  price
   1:    1    1  1  1   1  0    0      1      0          0      0      0 451.26
   2:    2    1  1  1   2  1    0     -1      0          0      0      0 491.26
   3:    3    2  1  2   1  1    0      0      1          0      0      0 491.26
   4:    4    2  1  2   2  0    0      0     -1          0      0      0 491.26
   5:    5    3  1  3   1  1    0      0      0          0      0      0 491.26
  ---
8929: 8929 3839 79 45   2  1    0      1      1          0      1      0 304.95
8930: 8930 3839 79 45   3  0    0     -1     -1         -1     -1     -1 296.95
8931: 8931 3840 79 46   1  0    0      1     -1         -1     -1     -1 267.95
8932: 8932 3840 79 46   2  1    0      1      1          0      1      0 304.95
8933: 8933 3840 79 46   3  0    0     -1      1          0      1      0 366.95
```

The first five columns contain indices for each row in the dataset (`row`), the observation (`obs`), the participant identifier (`id`), the choice task number for each participant (`cs`), and the alternative number in each choice task (`alt`). The dummy-coded variable `ch` indicates which alternative was chosen in each observation (i.e., in each `cs` and `id` combination). The dummy variable `none` identifies the no-choice option – every time a choice alternative constitutes a no-choice within the ACBC *Screening* stage. The columns starting with "`att_`" are attributes using effects coding. The first (second) number identifies the attribute (level). The last level is the reference level within each attribute. Hence, attributes 1 and 2 have two levels each, whereas attribute 7 has four levels (we omit attributes three to six to preserve horizontal space in Listing E2). The last column (`price`) indicates the price of each alternative (as the sum over all components, plus a random shock, see e.g., Johnson and Orme 2007). The none option has a price of zero, and all prices are scaled in €. However, to facilitate the Bayesian estimation, it is advisable to rescale variables to ensure that the resulting parameters have a similar magnitude close to unity.

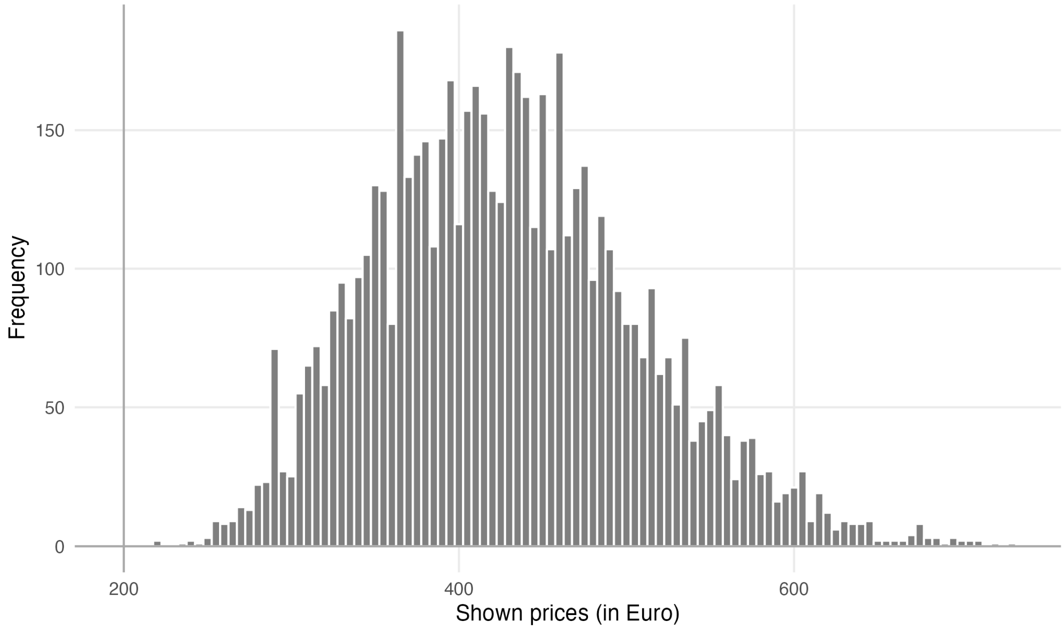


Fig. E2. Histogram of shown prices

Regarding the second data preparation step, as indicated in the article (and described in Web Appendix D), we de-mean prices and rescale them into €100 (in Supplemental Study A). However, Sawtooth maps prices to the [-1, 1]-interval, as we also do here, to benchmark everything as closely as possible.[6] Figure E2 depicts a histogram of the prices of all products (except for the none option). Prices have an average value of €428.99, starting at €220.95 and increasing to €730.95 (range = €510). After normalizing the prices using a simple linear transformation ($price_{norm} = \frac{(-(\max+\min) + 2 \cdot price)}{\max - \min} = -1.866 + 0.004 \cdot price$), the minimum, maximum, and range are -1, 1, and 2, respectively. This linear transformation does not affect the shape of the price distribution. Because the price distribution is not perfectly symmetric, the mean of the normalized prices is not zero but -0.184.[7] The function `fix_price_var()` performs this step (de-meaning or normalizing prices).

ACBC studies allow an optional fourth stage; the none parameter calibration. We also wrote an R-function for this task. The information for this step is not included in the *.cho file

[6] Linear transformations of the price only affect the estimation of the none and the price parameters.
[7] Note that this is not the case when de-meaning the prices, as we do in the main article. There, the mean of the transformed prices is, of course, zero.

but is included in the data exported from Sawtooth, which also contains participant-level information (e.g., age, gender, income, additional rating scales, etc.; saved in the file `data_ind.csv`). The relevant data parts consist of information on the products, which were additionally shown to the participant, as well as their answers on rating scales. The products must be constructed from stored information about the (individual) experimental designs.

Listing E3. Content and structure of reshaped calibration data

```
     row obs id cs alt cal none att_11 att_21 ... att_71 att_72 att_73  price
  1:   1   1  1  1   0   3    0     -1      1          0      1      0 491.26
  2:   2   2  1  2   1   3    0      1     -1          0      1      0 543.95
  3:   3   3  1  3  23   4    0     -1      1          0      1      0 481.95
  4:   4   4  1  4  31   4    0     -1      1          0      1      0 317.95
  5:   5   5  1  5   3   2    0     -1      1          0      1      0 691.95
 ---
468: 468 468 79  2   2   2    0     -1     -1         -1     -1     -1 478.95
469: 469 469 79  3   6   3    0      1     -1         -1     -1     -1 267.95
470: 470 470 79  4  27   2    0      1     -1         -1     -1     -1 307.95
471: 471 471 79  5   3   1    0      1      1          1      0      0 467.95
472: 472 472 79  6  18   4    0      1      1          0      1      0 304.95
```

Listing E3 shows the constructed calibration data after using the function `build_calibration_data()`. The structure is the same as the choice data. However, two points should be considered. First, instead of the `ch` column, the calibration data includes the column `cal` with participant ratings for each calibration alternative (i.e., a higher rating equates with a higher purchase likelihood). Second, prices are still scaled in € and must be transformed, as described above. We have to apply the minimum and maximum values for the prices from the choice data to be consistent across both data sources (`apply_price_fix()`).

Listing E4. Content and structure of reshaped hot data

```
       row obs id hot alt ch none att_11 att_21 ... att_71 att_72 att_73  price
   1:    1   1  1   1   1  0    0     -1     -1          0      0      1 427.65
   2:    2   1  1   1   2  0    0      1     -1          0      0      1 450.50
   3:    3   1  1   1   3  0    0      1     -1          0      0      1 437.49
   4:    4   1  1   1   4  0    0      1     -1          1      0      0 346.28
   5:    5   1  1   1   5  0    0     -1      1          0      1      0 511.50
  ---
1023: 1023  79 79   1   9  0    0     -1      1          1      0      0 498.75
1024: 1024  79 79   1  10  0    0      1      1          0      0      1 387.04
1025: 1025  79 79   1  11  0    0     -1      1         -1     -1     -1 482.06
1026: 1026  79 79   1  12  0    0      1      1         -1     -1     -1 418.86
1027: 1027  79 79   1  13  1    1      0      0          0      0      0   0.00
```

Listing E4 shows the hot data. Again, the structure is the same as the choice data. However, now we have the column `hot` instead of `cs`. Also, the prices need to be rescaled using `apply_price_fix()`.

Estimation

Table E1. Summary of estimation results (MNL and HB MNL)

	MNL (MLE)		HB MNL (HMC)			
	$\hat{b}$		$\hat{b}$		$\hat{\sigma}$	
Parameter	**Est**	**SE**	**Mean**	**95% CI**	**Mean**	**95% CI**
None	0.787	0.052	5.924	[5.070, 6.890]	2.470	[1.913, 3.083]
500GB	-0.174	0.035	-0.405	[-0.680, -0.159]	1.004	[0.743, 1.301]
Black	0.083	0.035	0.726	[0.372, 1.069]	1.371	[1.061, 1.752]
One dual shock controller	-0.232	0.036	-1.045	[-1.366, -0.740]	1.159	[0.894, 1.485]
Charging station	0.204	0.047	0.733	[0.447, 1.051]	1.006	[0.700, 1.336]
Wireless headset	-0.039	0.060	-0.334	[-0.655, -0.045]	0.880	[0.544, 1.245]
Far Cry Primal	0.099	0.065	0.346	[-0.028, 0.681]	1.190	[0.861, 1.554]
GTA V	0.296	0.061	0.771	[0.345, 1.179]	1.604	[1.219, 2.011]
Life is Strange	-0.141	0.070	-0.591	[-1.051, -0.150]	1.522	[1.136, 1.964]
TC's The Division	0.071	0.068	0.188	[-0.197, 0.598]	1.314	[0.941, 1.742]
Witcher 3	0.110	0.066	0.220	[-0.184, 0.628]	1.382	[0.994, 1.810]
Fallout 4	0.094	0.066	0.399	[0.022, 0.760]	1.197	[0.852, 1.599]
Final Fantasy x/x2	-0.115	0.072	-0.509	[-0.961, -0.081]	1.551	[1.172, 2.003]
Dark Souls 3	0.055	0.073	0.045	[-0.395, 0.480]	1.404	[0.980, 1.884]
Just Dance 2016	-0.504	0.075	-1.484	[-2.182, -0.860]	2.141	[1.570, 2.810]
Guitar Hero	0.083	0.065	-0.104	[-0.745, 0.505]	2.156	[1.668, 2.757]
FIFA 16	0.439	0.057	1.147	[0.466, 1.812]	2.652	[2.131, 3.246]
Price	-1.971	0.123	-6.251	[-7.701, -4.848]	5.310	[4.299, 6.503]

Notes: MNL: $LL = -2856.022$, HB MNL: $LMD = -1540.219$, $\min(n_{eff}) = 956$, $\text{median}(n_{eff}) = 1389$, and $\max(\hat{R}) = 1.005$. As we do not use a sign constraint, the price parameter is directly interpretable.

Before running the HB estimations in Sawtooth and Stan, we verify the correct data preparation. To this end, we run simple Maximum Likelihood Estimations (MLE) for models without heterogeneity (i.e., MNL). This is very simple in R (we use `mlogit`, see Croissant (2020), wrapped in the function `run_mnl()`) and also possible in Sawtooth, although somewhat hidden in the "Interaction Search Tool" (Sawtooth Software 2020b).

Table E1 indicates the parameter estimates of the MNL model. In this respect, the key finding is that the estimation of the MNL model with R leads to the same parameters and log-likelihood value as in Sawtooth. This underlines that our own data preparation steps, including the price normalization, are correct and in line with Sawtooth. Table E1 also includes the results ($\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{\sigma}}$) for the estimated HB MNL model, using Stan (using the function `b_sigma_summary()`). The model follows our explanation in Web Appendix D, except that we did not impose a sign constraint for the price parameter.

In Stan, we used the same priors as explained in Web Appendix D: $\boldsymbol{b} \sim N(0, 5)$, $\boldsymbol{\sigma} \sim N_{+}(0, 2.5)$, and $\boldsymbol{\Omega} \sim LKJCorr(2)$. We again used five chains, but in this case, with a reduced number of draws (1,700 per chain, with 200 draws as a warm-up and a thinning of five).[8] In Sawtooth (see, Sawtooth Software 2019a, p. 475), we employed the default setting for the priors (variance = 100 for the mean vector, as well as a variance of 1 and the degree of freedom equals the number of parameters in $\boldsymbol{b}$ plus five for the covariance matrix).

Because the regular Metropolis-Hastings algorithm is less efficient (i.e., slower convergence and more auto-correlation) than HMC, we used 200,000 draws in Sawtooth, where the first 120,000 were discarded as "warm-up" and every fifth draw was retained. This led to 16,000 draws from a single chain.

A comparison of the results in Table E1 shows that both models (with and without heterogeneity) produce similar, face-valid results. For example, the mean of the price parameter is negative. Participants also prefer (on average) a larger over a smaller hard disk (reference level = 1TB). Furthermore, we observe the typical result that the HB model's parameters have a larger scale, as the model fits much better compared to the homogenous MNL model. A comparison of the HB estimates in Table E1 and the results in Table D6 (with sign constraint) shows that they are nearly identical. The results for the price and the none parameters are slightly different, keeping in mind the different approaches to the rescaling of the prices.

[8] We believe that for the purpose of benchmarking, 1,500 posterior draws are sufficient. For the article's main analyses, we wanted to be on the "safe side" and aimed for a higher number of draws. Also, note that the number of parameters in Studies 1, 2, and 3 is slightly higher, and we used a setup that worked uniformly across all studies.

However, the log marginal density (LMD) value is higher (i.e., closer to zero) for the model with sign constraint (-1529.604 vs. -1540.219), implying a better fit.

The `sampling()`-function is part of the `rstan` package (Stan Development Team 2020a) and produces standardized output that is not only usable with our custom functions `b_sigma_summary()` or `get_posterior_means()`, but also in the whole "ecosystem" surrounding Stan in the R community. This is a great advantage because additional functionalities that work with the output from the `sampling()`-function are directly applicable to the estimated (A)CBC models.

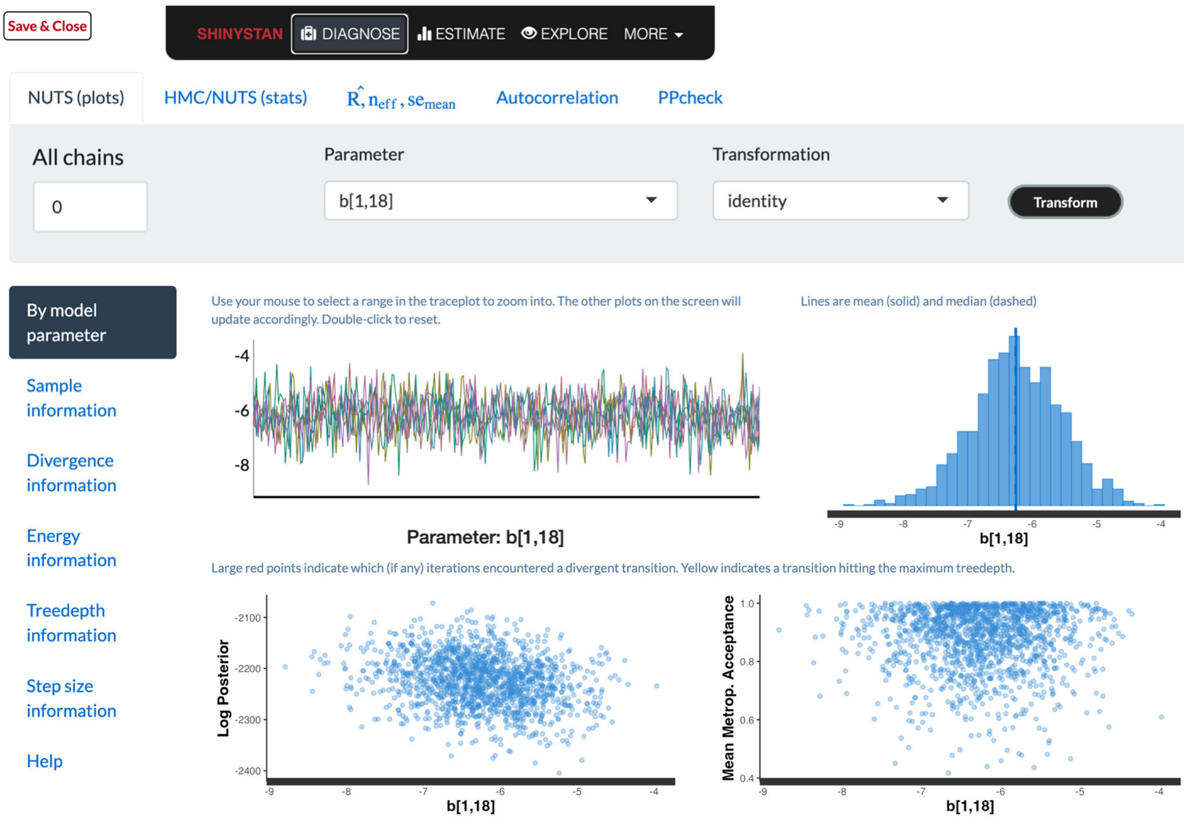


Fig. E3. Example output for `shinystan`

Besides functions for plotting in the R package `bayesplot` (Gabry and Mahr 2022), we want to mention and highlight the `shinystan` (Stan Development Team 2022). This package includes an interface to a dynamic Shiny dashboard (Chang et al. 2022), providing visual and numerical summaries for parameters as well as convergence diagnostics for the estimation. This is highly informative for the user and exceeds the reporting in Sawtooth by far. The user only needs to run the function `launch_shinystan()` on the output form `sampling()` to start the dynamic visualization and analysis GUI (see Fig. E1). Fig. E3 shows an example for the "DIAGNOSE" tab using the posterior draws of the model parameter $\hat{b}^p$ (see also Table E1). The overview depicts the traceplots for each chain, a histogram summarizing all information, as well as scatterplots for diagnostics. The results show, that the chains converged well, there are no signs of autocorrelation, the resulting parameter distribution looks symmetric with a normal shape, and there is no indication of divergent transitions.

Comparison

The tutorial thus far focused on the R and Stan results, respectively. To compare Stan and Sawtooth, it is necessary to consider the posterior mean estimates for each participant (i.e., $\widehat{\boldsymbol{\beta}}_i$) in more detail, which can be obtained by using the `get_posterior_means()`-function (see Fig. E1). Thus, we applied the calibration step in Sawtooth and R (based on the Stan results). Because we only briefly summarize the main points, the interested reader is referred to the manual for all details regarding this step (Sawtooth Software 2014).

Listing E5 shows the main R-function that is used for the calibration step (in `apply_calibration()` in Fig. E1). For each participant, we calculated the utility for the products in the calibration tasks using the participant's $\widehat{\boldsymbol{\beta}}_i$ parameters (input `x` of the function).

We then separately regressed the calibration ratings (from "definitely will not buy" [1] to "definitely will buy" [5], input `y` of the function and the `cal` column in Listing E3) on these utilities for each participant.[9] We further defined a threshold of 3 (which was the midpoint – "might or might not buy" – of the rating scale) and "calibrated" the none parameter by computing $\hat{\beta}_i^{none} = (\text{threshold} - \widehat{\text{intercept}}_i)/\widehat{\text{slope}}_i$, where $\widehat{\text{intercept}}_i$ and $\widehat{\text{slope}}_i$ represent the estimates from the aforementioned regression. Note that the calibration step is optional. In

[9] We implemented the five verbal scale anchors proposed by Rohrmann (1978) for probability judgments, as these German anchors were shown to be mentally represented equidistantly for participants.

the case of omitting it (in ACBC with 3 stages or CBC), we would directly transform the posterior means () or continue with predictions.

Listing E5. R-code for the calibration procedure

```
calibration = function(y, x, threshold = 3) {

  res = lm(y ~ x)
  b0 = round(coef(res)[1], 6)
  b1 = round(coef(res)[2], 6)

  if (b1 < 0) {
    if (all(y < threshold)) {
      none = max(x) + 1
      exception = "b"
    } else if (all(y >= threshold)) {
      none = min(x) - 1
      exception = "c"
    } else if (any(y >= threshold)) {
      none = mean(x[y >= threshold]) - 1
      exception = "a"
    }
  } else if (b1 == 0 & var(y) == 0) {
    ay = mean(y)
    if (ay < threshold) {
      none = ay + 1
      exception = "d"
    } else {
      none = ay - 1
      exception = "e"
    }
  } else if (b1 <= 0.05) {
    b1 = 0.05
    none = (threshold - b0) / b1
    exception = "f"
  } else {
    none = (threshold - b0) / b1
    exception = "none"
  }

  return(none)
}
```

We subdivide our comparison into three parts, each part focusing on a different aspect (i.e., first two moments, whole distribution, correlations).

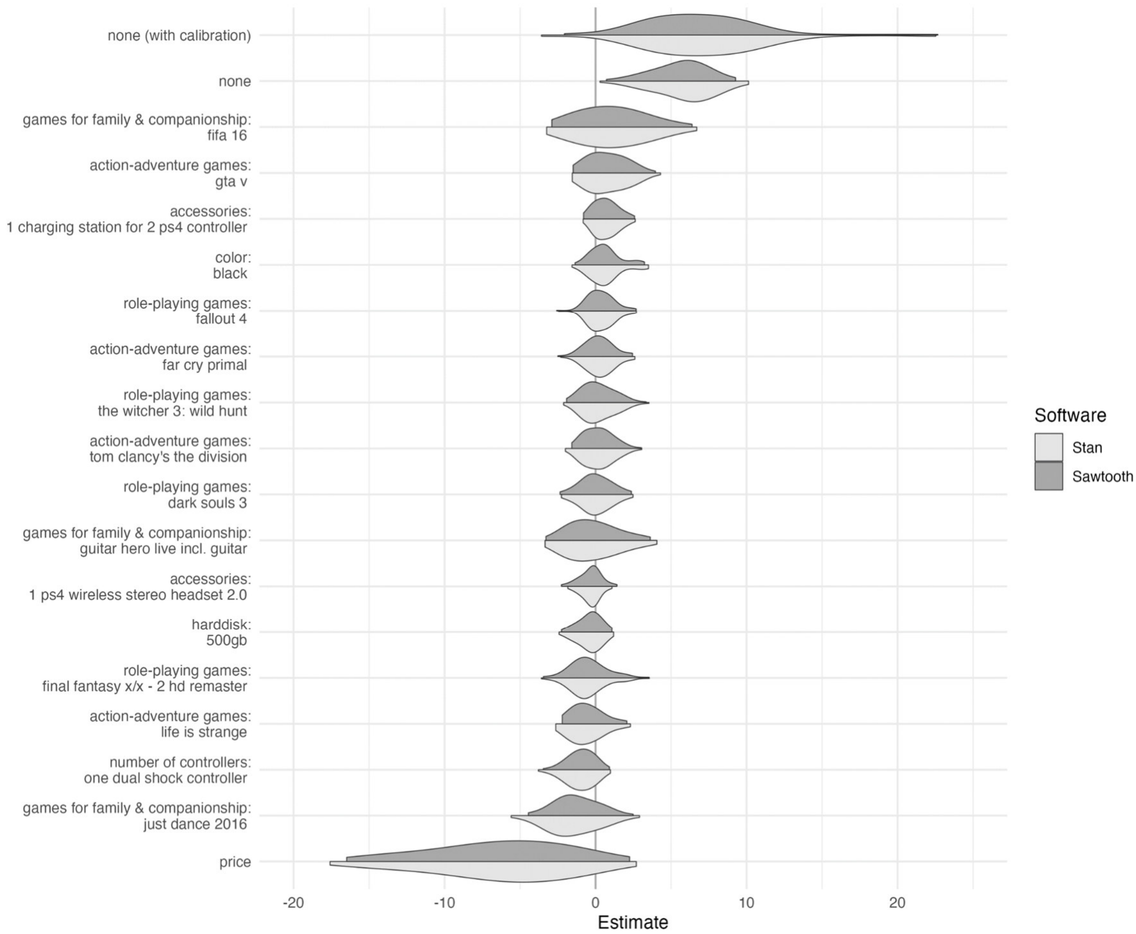
First, Table E2 summarizes the mean and standard deviation of the $\hat{\boldsymbol{\beta}}_i$ parameters from Stan and Sawtooth. For the Stan results, the user only needs to apply the function `partworth_summary()` on the output from the calibration (see Fig. E1). The results are nearly identical, with the same signs for all means and almost identical magnitudes for both means and standard deviations. A comparison of the values for both none parameters reveals that the calibrated version has larger values on average and contains more variation.

Table E2. Comparison of partworth estimates (Stan vs. Sawtooth)

	Stan		Sawtooth	
Parameter	**Mean**	***SD***	**Mean**	***SD***
None (3 stages)	5.959	2.022	5.595	1.780
None (4 stages)	7.713	5.818	7.310	5.704
500GB	-0.402	0.803	-0.381	0.741
Black	0.725	1.165	0.655	1.018
One dual shock controller	-1.054	0.966	-0.954	0.855
Charging station	0.734	0.760	0.673	0.776
Wireless headset	-0.333	0.602	-0.342	0.739
Far Cry Primal	0.337	0.939	0.234	0.955
GTA V	0.780	1.362	0.743	1.249
Life is Strange	-0.595	1.159	-0.557	1.040
TC's The Division	0.187	0.989	0.177	0.919
Witcher 3	0.220	1.073	0.162	1.035
Fallout 4	0.404	0.882	0.346	0.890
Final Fantasy x/x2	-0.503	1.236	-0.473	1.227
Dark Souls 3	0.041	1.050	0.027	1.021
Just Dance 2016	-1.466	1.701	-1.279	1.428
Guitar Hero	-0.109	1.807	-0.143	1.570
FIFA 16	1.156	2.444	1.034	2.195
Price	-6.364	4.749	-6.451	4.437

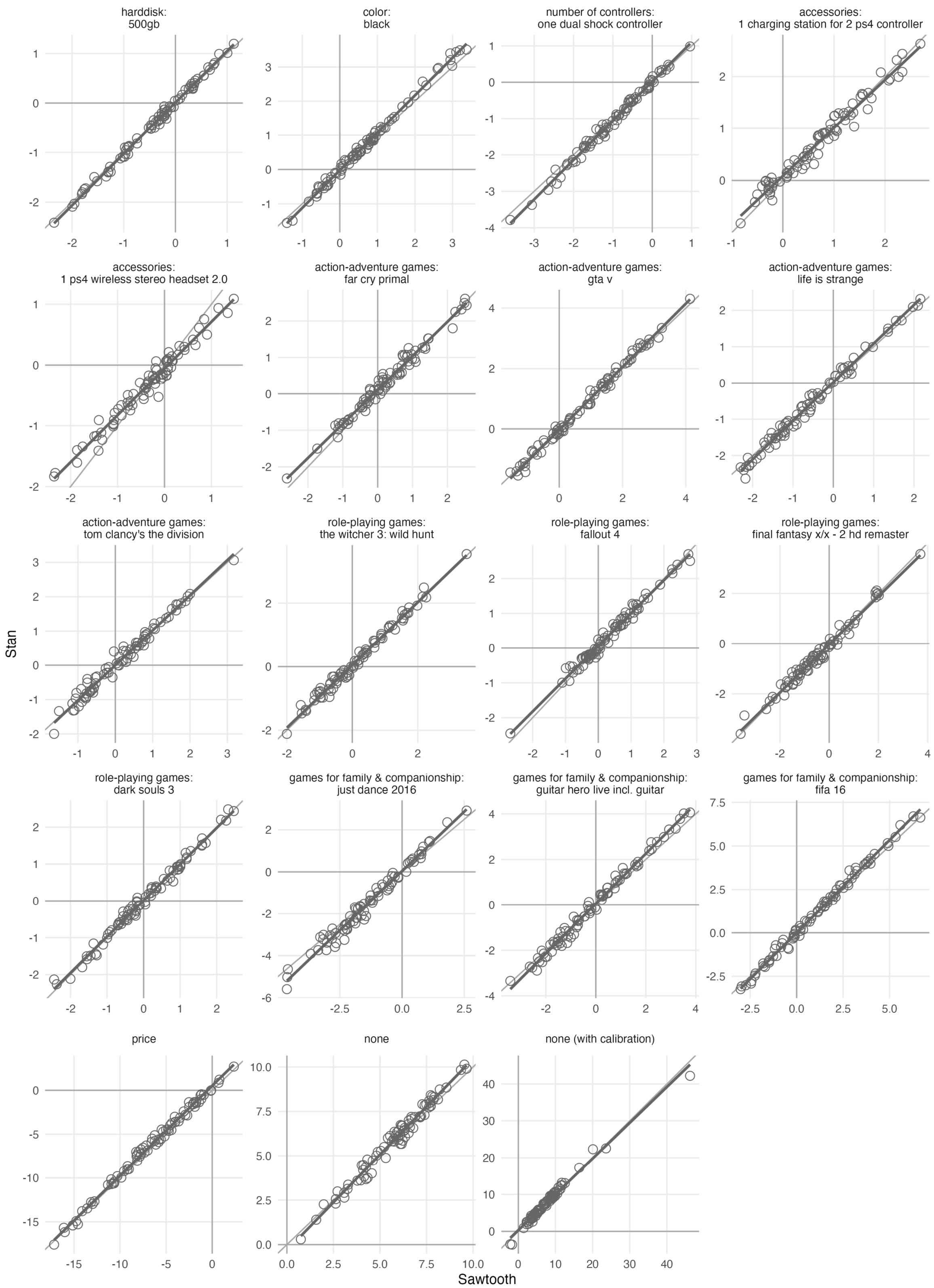
To examine the shapes of the whole $\widehat{\boldsymbol{\beta}}_i$ distributions, we also compared each parameter's densities using violin plots, depicted in Figure E3. The figure clearly shows that Stan and Sawtooth not only produce similar results in terms of the first two moments but also in respect of the whole shape of the $\widehat{\boldsymbol{\beta}}_i$ distributions. The differences are minimal and can be explained by the underlying uncertainty in the estimation.

Because we have two estimates for all participants and parameters – one estimate for each software – we can advance a step further in the last part and compare the results at the individual level. Figure E4 depicts the scatterplots for the participant-level parameters. A regression (without intercept) of the estimates from Stan on the estimates from Sawtooth reveals that the scale of the Stan parameters is slightly larger ($\approx +4\%$). Thus, Stan fitted slightly more towards individual data. For the scatterplot in Figure E4, we rescaled the Sawtooth estimates to ensure that both sets of estimates have the same scale (see Huber and Train (2001) for a similar approach when comparing results from Bayesian and Frequentist estimations). For each parameter, almost all points are tightly located close to the bisecting line. Also, for each parameter, the regression line lies almost perfectly on top of the bisecting line. Lastly, computing correlations with values close to 0.99 for all parameters shows that Stan and Sawtooth produce (almost) identical results at the participant level.



Notes: Parameters are ordered for a better interpretation by average values (i.e., smallest (largest) values at the bottom (top)). We also excluded a participant with a calibrated none parameter of > 40.

Fig. E3. Densities of partworth estimates



Notes: The dark (light) grey line is the regression (bisecting) line.

Fig. E4. Scatterplot of partworth estimates

Prediction

We implemented specific R functions to streamline and simplify the prediction based on the participant-level estimates of the models as well as the HOT evaluation. The function `predict_prob()` takes $\widehat{\boldsymbol{\beta}}_i$ parameters and HOT data as input and computes MNL probabilities (column `prob_hat`) for each participant (and HOT, in case data include multiple HOTs). Furthermore, based on the first-choice rule, predicted choices (column `ch_hat`) are computed.

$$\widehat{prob}_{ij}\left(\boldsymbol{x}_i, price_i, \widehat{\boldsymbol{\beta}}_i, \hat{\beta}_i^p\right) = \frac{\exp(\boldsymbol{x}_{ij}' \cdot \widehat{\boldsymbol{\beta}}_i - \hat{\beta}_i^p \cdot price_{ij})}{\sum_j^J \exp(\boldsymbol{x}_{ij'}' \cdot \widehat{\boldsymbol{\beta}}_i - \hat{\beta}_i^p \cdot price_{ij'})} \qquad \text{(formula E1)}$$

$$\widehat{ch}_{ij}\left(\boldsymbol{x}_i, price_i, \widehat{\boldsymbol{\beta}}_i, \hat{\beta}_i^p\right) = \begin{cases} 1, & if\ \widehat{prob}_{ij} = \max_{j' \in J}\left(\widehat{prob}_{ij'}\right) \\ 0, & else \end{cases} \qquad \text{(formula E2)}$$

Listing E6. Content and structure of predictions

```
       row obs id alt ch hot prob_hat condition stages ch_hat
   1:    1   1  1   1  0   1   0.0003  ACBC_inc      4      0
   2:    2   1  1   2  0   1   0.0001  ACBC_inc      4      0
   3:    3   1  1   3  0   1   0.0008  ACBC_inc      4      0
   4:    4   1  1   4  0   1   0.0103  ACBC_inc      4      0
   5:    5   1  1   5  0   1   0.0292  ACBC_inc      4      0
  ---
1023: 1023  79 79   9  0   1   0.0000  ACBC_inc      4      0
1024: 1024  79 79  10  0   1   0.0053  ACBC_inc      4      0
1025: 1025  79 79  11  0   1   0.0000  ACBC_inc      4      0
1026: 1026  79 79  12  0   1   0.0001  ACBC_inc      4      0
1027: 1027  79 79  13  1   1   0.9911  ACBC_inc      4      1
```

Listing E6 shows how the resulting object looks like. Once this object is created, out R function `compute_val_measures()` computes all relevant validation measures (hit rate, mean hit prob., Cohen's kappa, and mean absolute error at the aggregated level):

$$\mathrm{HR} = \frac{1}{I}\sum_{i=1}^{I}\sum_{j=1}^{J}\mathbb{I}(\widehat{ch}_{ij} = y_{ij}) \qquad \text{(formula E3)}$$

$$\mathrm{MHP} = \frac{1}{I}\sum_{i=1}^{I}\sum_{j=1}^{J}\widehat{prob}_{iy_{ij}} \qquad \text{(formula E4)}$$

$$\text{Cohen's kappa} = \frac{\mathrm{HR} - \mathrm{HR_e}}{1 - \mathrm{HR_e}} \qquad \text{(formula E5)}$$

$$\mathrm{MAE} = \frac{1}{J}\sum_{j=1}^{J}\left|share_j - \widehat{share}_j\right| \qquad \text{(formula E6)}$$

$\mathrm{HR_e}$ is the hypothetical probability of chance agreement, and $share_j$ and $\widehat{share}_j$ are the true and the predicted choice share for alternative j, respectively. In this case study (Supplemental Study A, incentive-aligned ACBC), we obtain a hit rate (in %) of 58.23, an MHP (in %) of 53.02, a Cohen's kappa of 0.43, and an MAE value of 1.65. These values agree well with the results reported in the main paper. However, as we dropped the sign constraint for the price parameter, the results are even more similar to the ones reported in Web Appendix G, Table G6 (Panel B, last column). As we reduced the number of MCMC draws to reduce estimation time and given the inherent randomness of MCMC methods, the results are not perfectly identical.

In conclusion, the tutorial shows that our implementation of the data processing, model estimation, and none parameter calibration leads to the same outcomes for all the analyses in the article, as would have been produced had we used Sawtooth. We hope that practitioners and academic researchers will adopt our code in future ACBC studies, because it is transparent and independent of Sawtooth Software and provides additional valuable features in conjoint research. Furthermore, modifications and extensions (e.g., estimation in WTP space, as suggested in Sonnier et al. 2007, but for ACBC data) would be straightforward to implement using our code.

Appendix F: Details regarding cost comparison of conjoint variants

For an estimation of project cost and time involved when conducting an incentive-aligned vs. hypothetical (A)CBC, we tasked two commercial European market research institutes that routinely handle conjoint studies to provide quotes for the study types we have conducted throughout the article (i.e., Studies 1-3 and Supplemental Study A).

The calculation (quotes are dated to 2019-2022) includes costs that are fixed per variant, independent of how many participants will be surveyed: setup cost, project management & programming, and analysis. Besides, it also includes variable costs that are a direct function of the number of participants surveyed: the panel fee, participants' fee, and cost for incentive alignment (i.e., products and shipping fees as study disbursement). All costs excluded value-added tax. Furthermore, we assume a case with n = 100 participants in each conjoint variant. We averaged across the two market research institutes, which gives the rounded cost estimates provided by Table F1.

Table F1. Elements of the cost function (€) for the conduction of different conjoint analysis variants with n = 100 participants

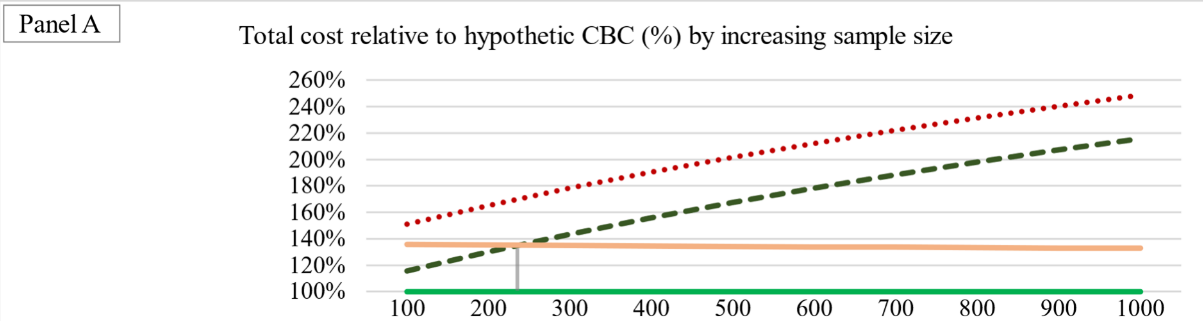
		Unit price in € / *Amount necessary per conjoint analysis variant (n = 100)*			
		Hypothetical CBC	Incentive-aligned CBC	Hypothetical ACBC	Incentive-aligned ACBC
	Project setup	200.00 / *1*	200.00 / *1*	200.00 / *1*	200.00 / *1*
	fee per participant	2.00 / *100*	2.00 / *100*	2.50 / *100*	2.50 / *100*
Incentive alignment	Pizza	-	15.00 / *100*	-	15.00 / *100*
	PS4	-	700.00 / *1*	-	700.00 / *1*
	Food processor		400.00 / *1*	-	400.00 / *1*
	Fitness tracker		640.00 / *1*	-	640.00 / *1*
	Management and programming	1,200 / *2*	1,200 / *2*	1,200 / *3*	1,200 / *3*
	analysis	1,200 / *2*	1,200 / *2*	1,200 / *2.5*	1,200 / *2.5*

Of course, these figures come with assumptions. The first assumption is that the participant fee is roughly equal, irrespective of the product category under research. The agencies assume for all four categories that it would not be a problem to recruit enough participants. However, in other studies, the relative incidence of eligible participants might be low (e.g., when running a conjoint study on specialized tools), potentially increasing the fee per participant. Second, the agencies point to the fact that every panel has a specific size and that, therefore, the fee per participant might increase if the total sample size of a study approaches the panel size. Third, as the additional time for the optional *Calibration* stage in ACBC is less than one minute, the agencies assume that the participant fees do not differ between ACBC with or without the *Calibration* stage.

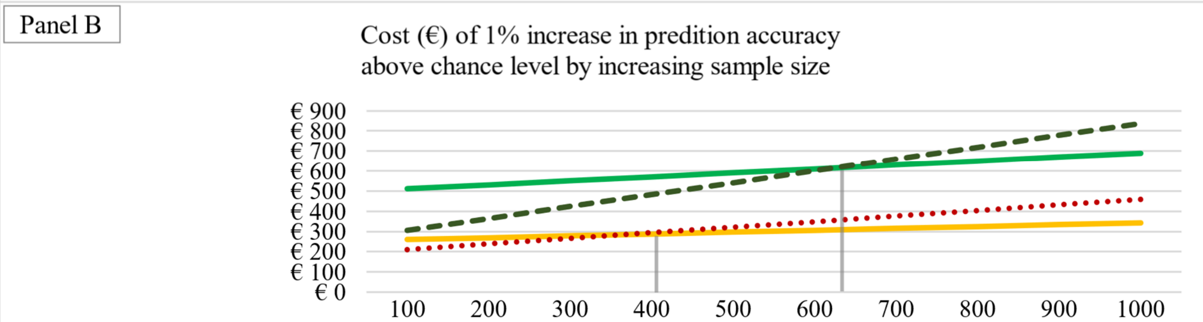
Based on the above figures, we first calculated the total cost per study variant (n = 100) as an average across the four product categories, which gives: €5,200 (hypothetical CBC), €6,010 (incentive-aligned CBC), €7,050 (hypothetical ACBC), and €7,860 (incentive-aligned ACBC). If one sets hypothetical CBC to 100%, the other variants amount to an additional total cost of 16% (incentive-aligned CBC), 36% (hypothetical ACBC), and 51% (incentive-aligned ACBC). Thus, in the frame of sample sizes that are comparable to the four presented studies, incentive-aligned ACBC has the highest absolute cost and hypothetical CBC has the lowest absolute cost.

This might, however, provide a biased picture, considering that commercial market research studies might consider much higher sample sizes to be representative in terms of diverse market segments (Natter et al. 2008; Desarbo et al. 1995) such as different countries etc. To shed light on the interplay between fixed and variable costs related to varying sample sizes, we conducted further sensitivity analysis. Specifically, we estimated the total cost of the different conjoint variants drawing on the quotes (Table F1), while systematically increasing the sample size.

In the underlying analysis, we calculate with a pizza voucher for each participant, whereas, for the food processor, the PS4, and the fitness tracker category, we assume that every hundredth participant will win a product. The results (Figure F1) deliver important insights.



	100	200	300	400	500	600	700	800	900	1000
hypothetic CBC	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
incentive-aligned CBC	116%	130%	143%	156%	168%	178%	189%	198%	207%	216%
hypothetic ACBC	136%	135%	135%	134%	134%	134%	134%	133%	133%	133%
incentive-aligned ACBC	151%	165%	178%	190%	202%	212%	222%	232%	240%	249%



	100	200	300	400	500	600	700	800	900	1000
hypothetical CBC	€ 512	€ 531	€ 551	€ 571	€ 590	€ 610	€ 630	€ 649	€ 669	€ 689
incentive-aligned CBC	€ 306	€ 365	€ 424	€ 483	€ 542	€ 601	€ 660	€ 719	€ 778	€ 837
hypothetical ACBC	€ 260	€ 269	€ 278	€ 288	€ 297	€ 306	€ 315	€ 324	€ 334	€ 343
incentive-aligned ACBC	€ 211	€ 239	€ 266	€ 294	€ 321	€ 349	€ 376	€ 404	€ 431	€ 459

Fig. F1. Sensitivity analysis of total cost and cost per prediction accuracy

Panel A of Figure F1 presents the total project cost (€) as a percentage compared to the hypothetical CBC (100%). First, one easily sees that – irrespective of the sample size – incentive-aligned ACBC remains the most expensive and hypothetical CBC the cheapest variant. The difference between both variants even widens with increasing sample sizes, due to the large variable cost for incentive alignment. Second, the ranking of conjoint variants regarding total cost changes between a sample size of n = 200 and n = 300. With smaller sample sizes, incentive-aligned CBC is cheaper than hypothetical ACBC. The latter variant comes with higher fixed cost. This ranking reverses for larger sample sizes because the variable cost of incentive alignment in incentive-aligned CBC then exceeds the fixed cost of hypothetical ACBC. Since our analysis of predictive validity in the manuscript highlights that both conjoint variants often had a head-to-head race, one can conclude that high sample sizes deliver an additional argument for preferring hypothetical ACBC instead of incentive-aligned CBC and vice versa.

In applied market research, the goal of conjoint analysis is usually to predict consumer purchase decisions. When focusing on the incentive-aligned HOTs implemented in our studies, one could guess which alternative a participant will choose, even without knowing that participant's utility function (i.e., without conducting a conjoint analysis). The accuracy of this bet is expected as one divided by the number of alternatives that a HOT provides, namely the chance level of prediction. As a logical corollary, the true benefit of the application of conjoint analysis is given by its ability to increase the quality of forecasts above the chance level.

Against this background, we next calculated how many percentage points each conjoint variant was able to increase the hit rate above the chance level (averaged across product categories and different HOTs within Study 3). Hypothetical CBC provides the poorest enhancement (10.16%), followed by incentive-aligned CBC (23.68%), hypothetical ACBC (29.20%), and finally, incentive-aligned ACBC (39.08%).

Finally, a marketing research manager might ask, how much money one must take in hand to "purchase" one percentage point prediction accuracy above the chance level in a market research study. The answer is given by the division of total cost by the increase in hit rate reported before and depends on the conjoint variant: €511.68 (hypothetical CBC), €305.64 (incentive-aligned CBC), €259.88 (hypothetical ACBC), and €211.19 (incentive-aligned ACBC). In contrast to the absolute cost, these figures speak a different language. Specifically, with sample sizes comparable to our four studies, both hypothetical and incentive-aligned

ACBC are linked to much lower cost per prediction accuracy, as compared to their CBC counterparts.

We also present a sensitivity analysis for this measure in Figure F1 (Panel B). Thereby, we assume that increasing sample size will not automatically increase hit rates (i.e., after all, with our initial sample sizes, we found stable utility estimates). The learnings extracted from this analysis are three-fold, as the initial ranking of conjoint variants changed completely with increasing sample size. First, for sample sizes larger than $n = 400$, the cost for one percent point prediction accuracy above the chance level in incentive-aligned ACBC exceeds those of hypothetical ACBC. Second, above a sample size of between $n = 600$ and $n = 700$, the costs of incentive-aligned CBC exceed those of the hypothetical CBC. In conclusion, in the absence of the ability to increase sample sizes to foster predictive validity substantially, there seems to be an upper limit for the relative advantageousness of incentive alignment in conjoint studies with very large sample sizes.

Third, due to their superior predictions per se, both ACBC variants remain superior over their CBC counterparts in terms of their lower cost per prediction accuracy. To summarize, when focusing on the total cost, the advantage of incentive-aligned CBC over hypothetical ACBC gets lost for sample sizes n > [200; 300]. Furthermore, when focusing on the cost per prediction accuracy, ACBC has an advantage over CBC, irrespective of sample sizes and incentive alignment or not. Finally, without the ability of incentive-aligned conjoint variants to further increase hit rates of prediction with increasing sample sizes, their advantage in cost-per-prediction accuracy can be compromised by large sample sizes.

Appendix G: Robustness analysis

Extreme response behavior

No-purchase option Other researchers conceptualize extreme response behavior as always or never choosing the no-purchase option in a CBC (Gensler et al. 2012; Schlereth and Skiera 2017). Data from participants exhibiting this behavior hamper conclusions about the maximum price that they would be willing to pay for a product concept. According to Gensler et al. (2012), even a small proportion of extreme responses in the dataset may affect the utility function estimates. These authors, therefore, recommend that CBC studies should report the share of participants who (1) always choose the no-purchase option and the share of those who (2) never choose the no-purchase option. Table G1 provides this information.

Table G1. Extreme response behavior in the CBCs

	Never		Always		Total	
Study	**count**	**%**	**count**	**%**	**count**	**%**
Study 1 (Pizza)						
Incentive-aligned CBC (n = 95)	10	10.53	8	8.42	18	18.95
Supplemental Study A (PS4)						
Incentive-aligned CBC (n = 84)	4	4.76	2	2.38	6	7.14
Study 2 (Food processor)						
Hypothetical CBC (n = 105)	37	35.24	3	2.86	40	38.10
Incentive-aligned CBC (n = 107)	40	37.38	5	7.48	45	42.06

Notes: Never = never chose the no-purchase option; Always = always chose the no-purchase option.

The share of extreme responses across studies ranged from 7.14% to 42.06%. These values are below or at the lower end of the range reported in previous studies (34.1–62.4%) (Schlereth and Skiera 2017; Mihale-Wilson et al. 2019; Ebbers et al. 2021; Halme and Kallio 2014).

Please note that Gensler et al. (2012) did not use adaptive designs following Johnson and Orme (2007). Web Appendix A outlines more details on their approach. In ACBC participants cannot opt for a no-purchase option in ACBC's *Choice Tournament* stage. Therefore, choice patterns in ACBC are not informative regarding extreme responses.

Table G2. Predictive validity of conjoint methods after excluding extreme responses

HOT prediction (number of products, format)	Hypothetical CBC		Incentive-aligned CBC	
	Full sample	No "extreme" responses	Full sample	No "extreme" responses
Panel A: Study 1 (16, forced-choice)			$n = 95$	$n = 77$
Hit rate in %			22.11	25.97
MHP in %			17.15	17.37
Cohen's kappa			0.15	0.19
MAE in %			2.38	2.83
Panel B: Supplemental Study A (12, free-choice)			$n = 84$	$n = 78$
Hit rate in %			48.81	46.15
MHP in %			35.07	32.92
Cohen's kappa			0.40	0.27
MAE in %			3.69	3.94
Panel C: Study 2	$n = 105$	$n = 65$	$n = 107$	$n = 62$
HOT 1 (10, forced-choice)				
Hit rate in %	19.05	16.92	26.17	32.26
MHP in %	16.94	16.37	19.05	22.06
Cohen's kappa	0.05	0.05	0.12	0.21
MAE in % (valid. sample)	3.70	4.84	4.88	4.08
HOT 2 (6, forced-choice)				
Hit rate in %	26.71	29.23	33.64	27.23
MHP in %	24.68	25.13	27.98	27.83
Cohen's kappa	0.06	0.09	0.17	0.08
MAE in % (valid. sample)	3.33	4.28	2.91	4.49
HOT 3 (6, free-choice)				
Hit rate in %	22.86	32.31	37.38	45.16
MHP in %	23.51	26.98	32.23	34.78
Cohen's kappa	0.06	0.14	0.22	0.27
MAE in % (valid. sample)	9.22	7.42	6.87	3.97
HOT 4 (6, ranking)				
Avg. rank of pred. choice	2.92	2.95	2.87	2.87
Avg. rank correlation	0.29	0.28	0.31	0.29
HOT 5 (4, forced-choice)				
MAE in % (valid. sample)	16.02	19.94	13.00	14.24
HOT 6 (4, free-choice)				
MAE in % (valid. sample)	14.80	10.53	15.03	9.43

Notes: Green colored entries = improvement of the predictive validity metric after excluding extreme responses; red-colored entries = deterioration of the predictive validity metric after excluding extreme responses.

Next, we checked whether and to what extent extreme responses impacted the predictive validity of the CBCs. Table G2 compares the results on the predictive validity metrics obtained for the full sample with those obtained for the sample that does not contain participants who gave extreme responses.

As is evident from Table G2, excluding extreme responses produced mixed results across and within studies. While Study 1, which involved a forced-choice HOT, showed that three out of four predictive validity metrics improved after excluding extreme responses, for Supplemental Study A, which included a free-choice HOT, the predictive validity metrics deteriorated. Turning to Study 2, which implemented a series of HOTs, several observations can be made. After excluding of extreme responses, the predictive validity in the (forced-choice) HOT 1 decreased for the hypothetical CBC and increased for the incentive-aligned CBC, while the opposite was true in (forced-choice) HOT 2. For the (free-choice) HOTs 3 and 6, excluding extreme responses improved the predictive validity of both CBC variants but impaired it in the (forced-choice) HOT 5 and in the (ranking) HOT 4. Although the results from Study 2 suggest that the predictive validity, as measured by the free-choice HOTs, benefits from excluding extreme responses, the results of Supplemental Study A contradict this notion.

Finally, we checked whether the effects of incentive alignment and adaptive designs on predictive validity metrics (e.g., hit rates and MHP) are robust to removing extreme responses. Table G3 compares the outcomes from the generalized mixed-effects models, ordered logistic regressions, and linear regressions that included extreme responses with those that did not.

The model estimates suggest that most of the effects that were found statistically significant in the analyses of all responses remained significant after excluding extreme responses. The effects that became non-significant still display a trend ($p < 0.10$) toward significance. The loss of significance for some effects after excluding extreme responses is not only attributable to changes in effect sizes, but also to the reduction in sample size. This was especially the case for Study 2, in which the sample size decreased from 105 to 65 in the hypothetical CBC and from 107 to 62 in the incentive-aligned CBC.

Table G3. Statistical analysis of predictive validity of conjoint methods before and after excluding extreme responses

	Full sample	No extreme responses
Panel A: Study 1		
Hit rate (logistic regression)		
hACBC vs. iCBC	$\beta = 0.57$, $z = 1.71$, $p = 0.087$	$\beta = 0.35$, $z = 1.04$, $p = 0.298$
iACBC vs. iCBC	$\beta = 1.22$, $z = 3.74$, $p_{(\text{one-tailed})} < 0.001$	$\beta = 1.00$, $z = 3.00$, $p_{(\text{one-tailed})} = 0.001$
MHP (linear regression)		
hACBC vs. iCBC	$\beta = 0.20$, $t_{(275)} = 1.42$, $p = 0.157$	$\beta = 0.242$, $t_{(257)} = 1.629$, $p = 0.105$
iACBC vs. iCBC	$\beta = 0.64$, $t_{(275)} = 4.46$, $p_{(\text{one-tailed})} < 0.001$	$\beta = 0.66$, $t_{(257)} = 4.40$, $p_{(\text{one-tailed})} < 0.001$
Panel B: Supplemental Study A		
Hit rate (logistic regression)		
hACBC vs. iCBC	$\beta = -0.18$, $z = -0.58$, $p = 0.565$	$\beta = -0.075$, $z = -0.233$, $p = 0.816$
iACBC vs. iCBC	$\beta = 0.33$, $z = 1.04$, $p_{(\text{one-tailed})} = 0.149$	$\beta = 0.43$, $z = 1.35$, $p_{(\text{one-tailed})} = 0.088$
MHP (linear regression)		
hACBC vs. iCBC	$\beta = 0.12$, $t_{(239)} = 0.77$, $p = 0.444$	$\beta = 0.17$, $t_{(233)} = 1.07$, $p = 0.288$
iACBC vs. iCBC	$\beta = 0.40$, $t_{(239)} = 2.59$, $p_{(\text{one-tailed})} = 0.005$	$\beta = 0.45$, $t_{(233)} = 2.85$, $p_{(\text{one-tailed})} = 0.002$
Panel C: Study 2		
HOTs 1-3		
Hit rate (ME logistic regression)		
Main effect – adaptive designs	$\beta = 0.64$, $z = 5.25$, $p_{(\text{one-tailed})} < 0.001$	$\beta = 0.49$, $z = 3.49$, $p_{(\text{one-tailed})} < 0.001$
Main effect – incent. alignment	$\beta = 0.41$, $z = 3.37$, $p_{(\text{one-tailed})} < 0.001$	$\beta = 0.37$, $z = 2.73$, $p_{(\text{one-tailed})} = 0.003$
hACBC vs. iCBC	$\beta = 0.23$, $z = 1.40$, $p = 0.162$	$\beta = 0.13$, $z = 0.66$, $p = 0.508$
MHP (ME linear regression)		
Main effect – adaptive designs	$\beta = 0.18$, $t_{(419)} = 2.88$, $p_{(\text{one-tailed})} = 0.002$	$\beta = 0.16$ $t_{(334)} = 2.25$, $p_{(\text{one-tailed})} = 0.013$
Main effect – incent. alignment	$\beta = 0.14$, $t_{(419)} = 2.24$, $p_{(\text{one-tailed})} = 0.013$	$\beta = 0.12$, $t_{(334)} = 1.74$, $p_{(\text{one-tailed})} = 0.041$
hACBC vs. iCBC	$\beta = 0.04$, $t_{(419)} = 0.46$, $p = 0.644$	$\beta = 0.04$, $t_{(334)} = 0.40$, $p = 0.687$
HOT 4 (ranking)		
Ranks of predicted choice (ordered logistic regression)		
Main effect – adaptive designs	$\beta = 0.55$, $z = 3.12$, $p_{(\text{one-tailed})} < 0.001$	$\beta = 0.47$, $z = 2.31$, $p_{(\text{one-tailed})} = 0.010$
Main effect – incent. alignment	$\beta = 0.29$, $z = 1.66$, $p_{(\text{one-tailed})} = 0.049$	$\beta = 0.31$, $z = 1.56$, $p_{(\text{one-tailed})} = 0.060$
hACBC vs. iCBC	$\beta = 0.26$, $z = 1.04$, $p = 0.297$	$\beta = 0.17$, $z = 0.59$, $p = 0.556$
Rank correlations (Spearman, linear regression)		
Main effect – adaptive designs	$\beta = 0.22$, $t_{(418)} = 2.30$, $p_{(\text{one-tailed})} = 0.011$	$\beta = 0.20$, $t_{(333)} = 1.83$, $p_{(\text{one-tailed})} = 0.034$
Main effect – incent. alignment	$\beta = 0.05$, $t_{(418)} = 0.52$, $p_{(\text{one-tailed})} = 0.302$	$\beta = 0.01$, $t_{(333)} = 0.07$, $p_{(\text{one-tailed})} = 0.472$
hACBC vs. iCBC	$\beta = 0.17$, $t_{(419)} = 1.28$, $p = 0.202$	$\beta = 0.20$, $t_{(334)} = 1.27$, $p = 0.204$
iACBC vs. hCBC	$\beta = 0.27$, $t_{(419)} = 1.96$, $p_{(\text{one-tailed})} = 0.025$	$\beta = 0.21$, $t_{(334)} = 1.36$, $p_{(\text{one-tailed})} = 0.088$

Notes: ME = mixed-effects; red-colored entries indicate that the effect was no longer significant after excluding extreme responses. The interaction effect between adaptive designs and incentive alignment remained statistically non-significant in all models. See the manuscript for more details.

Overall, the above analyses suggest that our findings regarding the predictive validity of different conjoint methods and the conclusions derived from them are largely robust to the exclusion of extreme responses. Since the main argument of Gensler et al. (2012) for excluding extreme responses is their negative impact on the validity of the WTP, we also compared reservation prices before and after excluding extreme responses. We found that the exclusion of extreme responses resulted in a more restricted range of reservation prices and a decrease in their medians. Despite this, the order of the aggregate reservation prices of ACBC and CBC conditions remained unaffected. More detailed results can be found in the subsection *Reservation prices* of Web Appendix H.

Random choice Another interpretation of "extreme response behavior" is choosing randomly. Contrary to extreme response behavior regarding the outside good, which mainly affects the none parameter, random choice behavior inflates the error term variance and consequently attenuates all parameter estimates. Howell et al. (2021) call such bad participants *"Gremlins,"* where *"the noise in their responses overwhelms any signal"* (p. 74).

To identify such Gremlins in the data, Howell et al. (2021) propose an advanced Bayesian model for CBC data that clusters Gremlins during the estimation. We followed Orme (2019) and employed a simpler approach that is similar to Howell et al. (2021) but relies on root likelihood (RLH) values (as a proxy for within-participant choice consistency) and derives a cutoff from a sample of random responses. To the best of our knowledge, we are the first to apply this approach to ACBC. Specifically, we simulate responses from 100 "fake" participants for the CBC and ACBC conditions using the setups of our four studies. Note that for the simulated participants, we do not have to distinguish between hypothetical and incentive-aligned conditions, but the difference between CBC and ACBC matters. Indeed, later-stage ACBC choice sets depend on (random) answers from previous stages. Hence, it is crucial to draw new answers to the ACBC choice sets and also build new ACBC datasets.

Next, we estimated the corresponding hierarchical Bayes models, drawing on our R/Stan implementation (Web Appendix D). Then, we calculated the RLH value for each random participant as an indication of individual-level internal model fit. For each dataset, we then took the upper limit of the 95% percentile as a measure of how high internal fit could be when only random answers are provided. Finally, we searched for individuals in our studies' real datasets with lower RLH values than the simulated thresholds.

Table G4. Analysis of random response behavior

	Random data RLH		**Real data RLH**	
Study	**Geom. Mean**	**Upper 95%**	**Minimum**	**Geom. Mean**
Study 1				
iCBC	0.271	0.297	0.484 (100%)	0.777
hACBC	0.446	0.475	0.583 (100%)	0.745
iACBC			0.607 (100%)	0.749
Supplemental Study A				
hCBC	0.263	0.281	0.469 (100%)	0.732
hACBC	0.463	0.509	0.613 (100%)	0.757
iACBC			0.585 (100%)	0.773
Study 2				
hCBC	0.265	0.290	0.315 (100%)	0.598
iCBC			0.246 (99.1%)	0.656
hACBC	0.456	0.487	0.523 (100%)	0.669
iACBC			0.518 (100%)	0.686
Study 3				
hACBC	0.471	0.508	0.566 (100%)	0.719
iACBC			0.629 (100%)	0.762

Notes: Values in parathesis are the fraction of participants with a minimum RLH value *above* the threshold of the 95%-percentile from the random data.

Table G4 provides an overview of the results. As expected, the (geometric) mean as a summary of in-sample fit is much lower for random data than for our real datasets. This is reassuring and shows that there is meaningful information to infer from the data using our models. Interestingly, in the random data, ACBC conditions still fit higher than CBC conditions. This is likely because some choice sets in ACBC have fewer options. This difference is less pronounced in the real data but still visible. Except for Study 1, the factors of incentive alignment and/or adaptive designs improve RLH values. Most importantly, the table allows for the comparison of the 95% percentile in RLH for the random data with the minimum (i.e., 0%-percentile) of the real data to identify the Gremlins. We can only identify one single participant (Study 2, incentive-aligned CBC) out of 1258 participants in 12 datasets across four studies, with an RLH value below the threshold. Thus, we conclude that "Gremlins" were not an issue in our analysis, most likely because of the proper screening of participants (i.e., general product interest was a requirement) and well-crafted choice studies.

HB models without sign constraint for the price parameter

As noted in the manuscript, the individual partworths were estimated using hierarchical Bayes (HB) models with a sign constraint for the price parameter to eschew economically implausible parameters (Allenby et al. 2014). Without this step, one would not be able to calculate individual WTP or reservation prices, or one would have to exclude participants with implausible price parameters. In this section of the Web Appendix, we first examined whether the proportion of (economically plausible) negative price parameters differs between studies and conditions. Table G5 presents the results.

Table G5. Proportions of negative price parameters

Study	Hypothetical CBC	Incentive-aligned CBC	Hypothetical ACBC	Incentive-aligned ACBC
Study 1				
Population-level		0.896	0.972	0.930
Individual-level		0.947	1.000	0.967
Supplemental Study A				
Population-level		0.961	0.867	0.868
Individual-level		0.988	0.899	0.949
Study 2				
Population-level	0.768	0.764	0.836	0.863
Individual-level	0.790	0.860	0.873	0.910
Study 3				
Population-level			0.926	0.896
Individual-level			0.952	0.952

Table G5 highlights that across all studies and conditions, individual-level models estimated a higher proportion of negative price parameters than the population-level models. This could indicate that the population-level models' normal assumption is overestimating the proportion of participants actually holding a positive price parameter (Allenby and Rossi 1998).

The comparison between CBCs and ACBCs showed no consistent pattern. While the proportion of negative price parameters is higher in the ACBCs in Studies 1 and 2, it is lower in Supplemental Study A. Similarly, there were no systematic differences between hypothetical and incentive-aligned ACBCs. However, Study 2 (the only study with all four conditions) suggests that adaptive designs and/or incentive alignment reduce wrong signs.

Next, we checked whether the results on the predictive validity are robust to using HB models without sign constraint for the price parameter. The results for the predictive validity metrics obtained from these models are presented in Table G6. The information presented in Table G6 indicates that the general pattern of results remained basically the same when comparing the results from HB models without a sign constraint to those from the HB models with a sign constraint. When considering each study individually, Study 1 was the only one in which the HB model without sign constraint provided (slightly) better values for all predictive validity metrics. In all other studies, no systematic differences were found between the two types of HB models. Overall, the findings regarding the advantage or disadvantage of one conjoint variant over another in terms of predictive validity are essentially unchanged, regardless of whether a sign constraint was imposed in the HB models.

Table G7 supports a similar conclusion. It compares the outcomes from various models on predictive validity metrics of conjoint methods using HB models with and without a sign constraint. Two contrasts reached statistical significance, and one effect became non-significant when the sign constraint was removed. Otherwise, no significant differences were observed between the two HB models.

The results show the robustness of our findings without imposing a sign constraint. HB models with or without the sign constraint are interchangeable for most analyses. However, imposing the sign constraint allows for a more economically sound analysis of WTP and reservation prices at the individual level, making it more suitable for this study.

Table G6. Predictive validity metrics of conjoint methods obtained from HB models without sign constraint

HOT prediction (number of products, format)	**Hypothetical CBC**	**Incentive-aligned CBC**	**Hypothetical ACBC**	**Incentive-aligned ACBC**
Panel A: Study 1 (16, forced-choice)		*n* = 95	*n* = 93	*n* = 90
Hit rate in %		23.16 (22.11)	34.41 (33.33)	50.00 (48.89)
MHP in %		17.46 (17.15)	24.54 (24.53)	35.33 (35.23)
Cohen's kappa		0.16 (0.15)	0.27 (0.26)	0.44 (0.43)
MAE in %		2.33 (2.38)	3.47 (3.51)	1.81 (1.83)
Panel B: Supplemental Study A (12, free-choice)		*n* = 84	*n* = 79	*n* = 79
Hit rate in %		50.00 (48.81)	45.57 (44.30)	59.49 (56.96)
MHP in %		35.45 (35.07)	43.89 (43.43)	53.25 (53.33)
Cohen's kappa		0.41 (0.40)	0.21 (0.20)	0.45 (0.42)
MAE in %		3.70 (3.69)	1.65 (1.65)	1.61 (1.52)
Panel C: Study 2	*n* = 105	*n* = 107	*n* = 110	*n* = 100
HOT 1 (10, forced-choice)				
Hit rate in %	20.00 (19.05)	25.23 (26.17)	36.36 (38.18)	40.00 (41.00)
MHP in %	16.63 (16.64)	19.91 (19.05)	27.96 (28.56)	32.84 (32.61)
Cohen's kappa	0.06 (0.05)	0.11 (0.12)	0.27 (0.29)	0.31 (0.32)
MAE in % (valid. sample)	3.88 (3.70)	5.02 (4.88)	4.36 (4.42)	3.43 (3.43)
HOT 2 (6, forced-choice)				
Hit rate in %	25.71 (25.71)	31.78 (33.64)	39.09 (39.09)	52.00 (50.00)
MHP in %	25.23 (24.68)	28.99 (27.98)	34.92 (35.14)	47.53 (47.22)
Cohen's kappa	0.06 (0.06)	0.15 (0.17)	0.26 (0.26)	0.40 (0.37)
MAE in % (valid. sample)	3.27 (3.33)	2.81 (2.91)	3.59 (3.39)	2.48 (2.54)
HOT 3 (6, free-choice)				
Hit rate in %	23.81 (22.86)	37.38 (37.38)	35.45 (35.45)	46.00 (46.00)
MHP in %	23.45 (23.51)	32.23 (32.46)	34.88 (34.62)	42.55 (42.50)
Cohen's kappa	0.07 (0.06)	0.22 (0.22)	0.13 (0.13)	0.22 (0.22)
MAE in % (valid. sample)	8.86 (9.22)	6.87 (7.08)	5.56 (5.58)	6.25 (6.40)
HOT 4 (6, ranking)				
Avg. rank of pred. choice	2.78 (2.92)	2.87 (2.87)	2.61 (2.55)	2.14 (2.30)
Avg. rank correlation	0.30 (0.29)	0.29 (0.31)	0.40 (0.42)	0.41 (0.41)
HOT 5 (4, forced-choice)				
MAE in % (valid. sample)	16.24 (16.02)	12.92 (13.00)	12.45 (12.44)	11.68 (11.45)
HOT 6 (4, free-choice)				
MAE in % (valid. sample)	14.38 (14.80)	14.75 (15.03)	9.47 (9.53)	10.23 (10.20)
Panel D: Study 3			*n* = 105	*n* = 104
(9, forced-choice)				
Hit rate in %			31.43 (36.19)	46.15 (45.19)
MHP in %			29.67 (29.77)	38.04 (38.17)
Cohen's kappa			0.16 (0.22)	0.29 (0.28)
MAE in %			5.91 (5.94)	3.47 (3.51)

Notes: Values in the parenthesis were obtained from HB models with sign constraint for the price (see Table 2 in the manuscript). Green colored entries = HB model without sign constraint performed better; red-colored entries = HB model without sign constraint performed worse.

Table G7. Statistical analysis of predictive validity of conjoint methods using estimates from HB models with and without sign constraint

	With sign constraint	**Without sign constraint**
Panel A: Study 1		
Hit rate (logistic regression)		
iACBC vs. hACBC	$\beta = 0.65$, $z = 2.13$, $p = 0.017$	$\beta = 0.65$, $z = 2.13$, $p = 0.017$
hACBC vs. iCBC	$\beta = 0.57$, $z = 1.71$, $p = 0.087$	$\beta = 0.55$, $z = 1.70$, $p = 0.090$
hACBC vs. iCBC	$\beta = 1.22$, $z = 3.74$, $p_{(\text{one-tailed})} < 0.001$	$\beta = 1.20$, $z = 3.73$, $p_{(\text{one-tailed})} < 0.001$
MHP (linear regression)		
iACBC vs. hACBC	$\beta = 0.43$, $t_{(275)} = 3.04$, $p_{(\text{one-tailed})} = 0.001$	$\beta = 0.43$, $t_{(275)} = 3.04$, $p_{(\text{one-tailed})} = 0.001$
hACBC vs. iCBC	$\beta = 0.20$, $t_{(275)} = 1.42$, $p = 0.157$	$\beta = 0.20$, $t_{(275)} = 1.45$, $p = 0.150$
iACBC vs. iCBC	$\beta = 0.64$, $t_{(275)} = 4.46$, $p_{(\text{one-tailed})} < 0.001$	$\beta = 0.64$, $t_{(275)} = 4.49$, $p_{(\text{one-tailed})} < 0.001$
Panel B: Supplemental Study A		
Hit rate (logistic regression)		
iACBC vs. hACBC	$\beta = 0.51$, $z = 1.59$, $p_{(\text{one-tailed})} = 0.056$	$\beta = 0.562$, $z = 1.75$, $p_{(\text{one-tailed})} = 0.040$
hACBC vs. iCBC	$\beta = -0.18$, $z = -0.58$, $p = 0.565$	$\beta = -0.18$, $z = -0.57$, $p = 0.572$
iACBC vs. iCBC	$\beta = 0.33$, $z = 1.04$, $p_{(\text{one-tailed})} = 0.149$	$\beta = 0.38$, $z = 1.22$, $p_{(\text{one-tailed})} = 0.112$
MHP (linear regression)		
iACBC vs. hACBC	$\beta = 0.28$, $t_{(239)} = 1.80$, $p_{(\text{one-tailed})} = 0.037$	$\beta = 0.26$, $t_{(239)} = 1.67$, $p_{(\text{one-tailed})} = 0.048$
hACBC vs. iCBC	$\beta = 0.12$, $t_{(239)} = 0.77$, $p = 0.444$	$\beta = 0.13$, $t_{(239)} = 0.85$, $p = 0.398$
iACBC vs. iCBC	$\beta = 0.40$, $t_{(239)} = 2.59$, $p_{(\text{one-tailed})} = 0.005$	$\beta = 0.40$, $t_{(239)} = 2.55$, $p_{(\text{one-tailed})} = 0.006$
Panel C: Study 2		
HOTs 1-3		
Hit rate (ME logistic regression)		
Main effect – adaptive designs	$\beta = 0.64$, $z = 5.25$, $p_{(\text{one-tailed})} < 0.001$	$\beta = 0.65$, $z = 5.18$, $p_{(\text{one-tailed})} < 0.001$
Main effect – incent. alignment	$\beta = 0.41$, $z = 3.37$, $p_{(\text{one-tailed})} < 0.001$	$\beta = 0.40$, $z = 3.22$, $p_{(\text{one-tailed})} = 0.001$
hACBC vs. iCBC	$\beta = 0.23$, $z = 1.40$, $p = 0.162$	$\beta = 0.25$, $z = 1.46$, $p = 0.143$
MHP (ME linear regression)		
Main effect – adaptive designs	$\beta = 0.18$, $t_{(419)} = 2.88$, $p_{(\text{one-tailed})} = 0.002$	$\beta = 0.18$ $t_{(419)} = 2.89$, $p_{(\text{one-tailed})} = 0.002$
Main effect – incent. alignment	$\beta = 0.14$, $t_{(419)} = 2.24$, $p_{(\text{one-tailed})} = 0.013$	$\beta = 0.14$, $t_{(419)} = 2.21$, $p_{(\text{one-tailed})} = 0.014$
hACBC vs. iCBC	$\beta = 0.04$, $t_{(419)} = 0.46$, $p = 0.644$	$\beta = 0.04$, $t_{(419)} = 0.49$, $p = 0.623$
HOT 4 (ranking)		
Ranks of predicted choice (ordered logistic regression)		
Main effect – adaptive designs	$\beta = 0.55$, $z = 3.12$, $p_{(\text{one-tailed})} < 0.001$	$\beta = 0.46$, $z = 2.62$, $p_{(\text{one-tailed})} = 0.004$
Main effect – incent. alignment	$\beta = 0.29$, $z = 1.66$, $p_{(\text{one-tailed})} = 0.049$	$\beta = 0.29$, $z = 1.64$, $p_{(\text{one-tailed})} = 0.050$
hACBC vs. iCBC	$\beta = 0.26$, $z = 1.04$, $p = 0.297$	$\beta = 0.17$, $z = 0.69$, $p = 0.491$
Rank correlations (Spearman, linear regression)		
Main effect – adaptive designs	$\beta = 0.22$, $t_{(419)} = 2.30$, $p_{(\text{one-tailed})} = 0.011$	$\beta = 0.21$, $t_{(419)} = 2.20$, $p_{(\text{one-tailed})} = 0.014$
Main effect – incent. alignment	$\beta = 0.05$, $t_{(419)} = 0.52$, $p_{(\text{one-tailed})} = 0.302$	$\beta = 0.03$, $t_{(419)} = 0.35$, $p_{(\text{one-tailed})} = 0.363$
hACBC vs. iCBC	$\beta = 0.17_{(419)}$, $t = 1.28$, $p = 0.202$	$\beta = 0.18_{(419)}$, $t = 1.33$, $p_{(\text{one-tailed})} = 0.185$
iACBC vs. hCBC	$\beta = 0.27$, $t_{(419)} = 1.96$, $p_{(\text{one-tailed})} = 0.025$	$\beta = 0.25$, $t_{(419)} = 1.78$, $p_{(\text{one-tailed})} = 0.038$
Panel D: Study 3		
Hit rate (logistic regression)		
iACBC vs. hACBC	$\beta = 0.37$, $z = 1.32$, $p_{(\text{one-tailed})} = 0.093$	$\beta = 0.63$, $z = 2.17$, $p_{(\text{one-tailed})} = 0.015$
MHP (linear regression)		
iACBC vs. hACBC	$\beta = 0.28$, $t_{(206)} = 2.05$, $p_{(\text{one-tailed})} = 0.021$	$\beta = 0.28$, $t_{(206)} = 2.00$, $p_{(\text{one-tailed})} = 0.023$

Notes: ME = mixed-effects; green colored entries indicate that the effect became significant when a HB model without sign constraint was used. Red-colored entry indicates that the effect was no longer significant when a HB model without sign constraint was used. The interaction effect between adaptive designs and incentive alignment remained statistically non-significant in all models. See the manuscript for more details.

Appendix H: Further analyses

Scale differences across conditions

In choice models, the absolute value of the deterministic part of the utility function and the error-term variance cannot be identified separately. Researchers using logit models, therefore, typically employ a specific version of the type I extreme value distribution for the error term, $\varepsilon_{ijt} \sim EV(0, \mu)$, and fix $\mu = 1$ (Sonnier et al. 2007). Consequently, utility parameters and the scale of the analysis $(1/\mu)$ are confounded and interpreting estimates across models and datasets is difficult because differences in estimates might come from differences in the error-term variance $(\mu^2\pi^2/6)$ and hence the scale. Recently, Hauser et al. (2019) pointed out that decisions such as using realistic images or incentive alignment in CBC studies can affect the scale. In this section, we analyze whether there are systematic differences in scale across conditions; differences in scale could, to some extent, explain our results regarding the predictive validity.[10] However, better predictions do not necessarily need to align with larger values of the scale (i.e., smaller error-term variance) and the larger scale just indicates that the model fit is better in-sample.

A simple way to assess scale differences is by comparing relative magnitudes of coefficients. Frischknecht et al. (2014) compare different estimation approaches using scatter plots and find that individual-level estimates are highly proportional, but hierarchical Bayesian estimation systematically leads to a smaller error variance and, thus, larger estimates in magnitude. Our study design forbids comparing individual estimates as each participant was randomly assigned to only one condition. Therefore, we compare mean partworths as they are equally affected by the scale. Specifically, we regress the partworths of the ACBC conditions with incentive alignment on conditions without alignment or without adaptive designs. This

[10] We thank one anonymous reviewer for bringing this topic to our attention.

allows us to analyze whether scale differences are present. The null hypothesis of no scale differences implies a slope of 1 in the regressions.[11]

Table H1. Results of scale regressions

Study	Comparison	Intercept (SE)	Slope (SE)	R^2
Incentive alignment				
Study 1 (Pizza)	iACBC vs. hACBC	0.001 (0.090)	**1.203 (0.061)**	0.941
Supplemental Study A (PS4)	iACBC vs. hACBC	0.056 (0.070)	0.989 (0.076)	0.919
Study 2 (Food Processors)	iACBC vs. hACBC	0.080 (0.037)	1.017 (0.051)	0.948
Study 3 (Fitness Trackers)	iACBC vs. hACBC	-0.101 (0.168)	1.004 (0.091)	0.938
Study 2 (Food Processors)	iCBC vs. hCBC	-0.014 (0.030)	**1.162 (0.075)**	0.916
Adaptive designs				
Study 1 (Pizza)	iACBC vs. iCBC	-0.115 (0.095)	**1.313 (0.072)**	0.933
Supplemental Study A (PS4)	iACBC vs. iCBC	-0.137 (0.108)	**1.467 (0.189)**	0.801
Study 2 (Food Processors)	iACBC vs. iCBC	-0.088 (0.051)	**1.552 (0.105)**	0.905
Study 2 (Food Processors)	hACBC vs. hCBC	**-0.186 (0.055)**	**1.743 (0.138)**	0.916

Notes: Significant parameters with $p < 0.05$ are highlighted in **bold** and slopes are tested against 1.

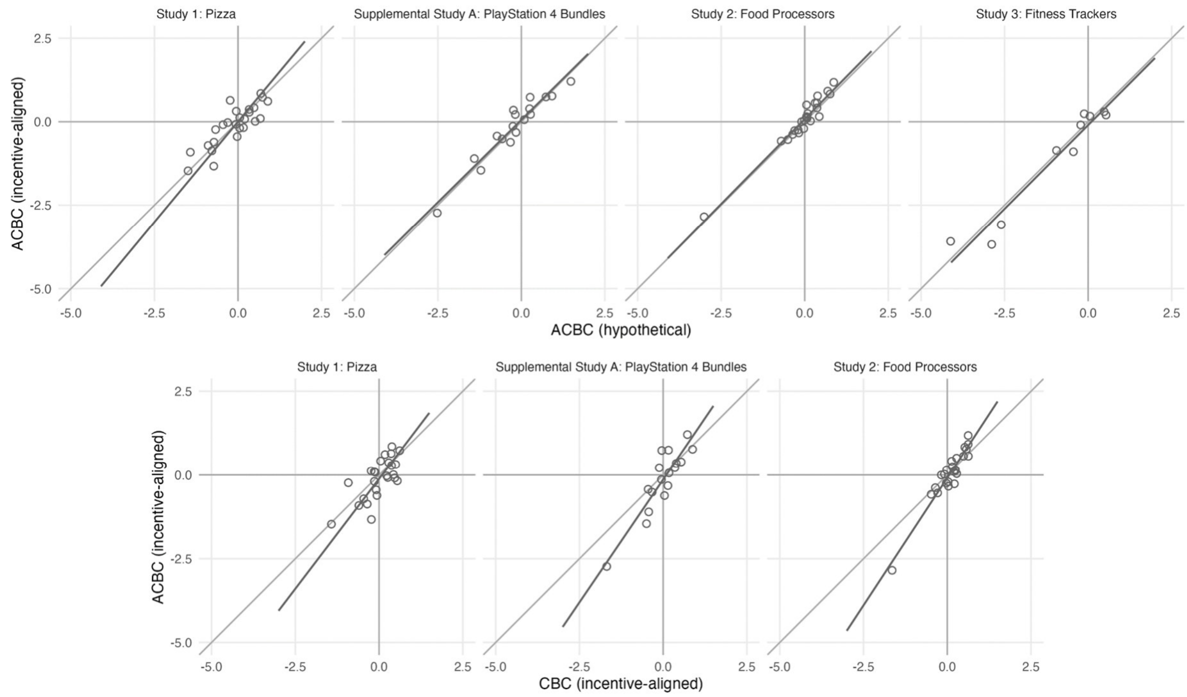


Fig. H1. Scatterplot of mean partworths

[11] An alternative approach would be a joint model and estimation for all conditions per study, such as in Hauser et al. (2019). This would enable us to assume the same population-level distribution for the heterogenous parameters. We have decided to apply a simpler version of the scale analysis using only mean partworths and leave the analysis based on a joint model and estimation for future research.

Table H1 summarizes the main regression results, and Figure H1 shows scatter plots for the specific comparisons, including the estimated regression lines for cases where incentive-aligned ACBC is shown on the *y*-axis. All regressions clearly show that the scales are similar (slopes close to 1, intercepts close to 0), and results are aligned across conditions (R^2 values of > 0.8). In three out of five cases, there are no differences in scale in conditions with and without incentive alignment. Only in Study 1 (ACBC comparison, $\beta = 1.203$, $t_{(24)} = 3.30$, $p = 0.003$) and Study 2 (CBC comparison: $\beta = 1.162$, $t_{(22)} = 2.16$, $p = 0.042$) the slope parameter is significantly larger than 1. Using a random effects single paper meta-analysis (McShane and Böckenholt 2017) to summarize the main effect of incentive alignment on the scale shows an effect of 1.079 (i.e., +8%), but the standard error of 0.046 implies no significant effect when testing against 1. Also, Cochran's Q-test for heterogeneity (Hedges and Olkin 1985) is not significant: $Q_{(4)} = 8.916$, $p = 0.063$. The results are less ambiguous when comparing conditions with and without adaptive designs. In four out of four cases (note that Study 3 has only ACBC conditions), we find significant increases with adaptive designs (vs. static CBC). The slope estimates are between 1.313 and 1.743, implying scale increases between 31% and 74%. This time, the meta-analytic effect is larger (1.502, i.e., about +50%) and statistically significant ($z = 5.002$, $p = 0.001$). We still do not find significant heterogeneity ($Q_{(3)} = 9.076$, $p = 0.028$). These results mainly support the conclusion that there are indeed significant scale differences when applying adaptive designs. Incentive alignment also increases the scale, but only descriptively (and very slightly in magnitude).

Based on these findings, we conclude that participants make more consistent (i.e., less random) choices in studies with adaptive designs. This is presumably driven by the more relevant options that the participants encounter during the study. The results align well with our findings regarding study enjoyment. On the other hand, we do not find strong evidence that higher deliberation through incentive alignment generally increases scale. Participants spend more time on the *BYO* section in the case of incentive alignment, but this does not (except for Study 1) improve choice consistency throughout the whole ACBC interview. This was expected to some extent, as Hauser et al. (2019) found that CBC studies with incentive alignment (vs. without) even have a smaller scale. In the end, market researchers should achieve results that align with participants' true (but unbeknown) scale. Given that our ACBC conditions outperform CBC conditions in terms of predictive validity, these results regarding scale are convincing. We do not see larger scales just because participants employ heuristics and simplification strategies that lead to a good fit of the model in-sample. Adaptive designs help to infer true preferences better, leading to a better (relative) performance in- and out-of-sample.

Reservation prices

Next, we broadened the scope of the investigation to marketing-related differences between methods apart from predictive validity. We now analyze reservation prices (see, e.g., Miller et al. 2011; Gensler et al. 2012; Jedidi and Zhang 2002; Wertenbroch and Skiera 2002):

$$rp_i = \left(u_{i|\sim p} - u_i^*\right)/\beta_i^p, \qquad \text{(formula H1)}$$

where $u_{i|\sim p}$ is the non-price utility, u_i^* the outside-good utility, and β_i^p the negatively defined price coefficient of participant i. This is i's maximum WTP for a specific product over the outside-good option (i.e., not purchasing the product). Compared to a simple rescaling of partworths in monetary space (WTP for features, see Web Appendix D), this measure takes the preference of the whole product into account and also considers the "competition" with the outside option.

To be able to compute rp_i in each condition (and study), we must select a product. For this purpose, we use a "popular" product in each study that most customers would buy. Table H2 briefly describes the specific attributes and levels we chose. Besides general popularity, we paid attention to the fact that, across conditions, the directions of preferences were similar (in Study 1, e.g., on average, participants prefer larger pizzas over smaller ones in all conditions).

Table H2. Product definitions for reference price analyses

Study	Product
Study 1	Pizza type: *Tomato*, Pizza size: *29 cm*, Additional sauces: *None*, Additional topping: *Paprika*, Additional cheese: *Mozzarella*, Additional spice: *None*, Beverages: *Coca-Cola*
Study 2	Color: *Black*, Power: *1000 Watt*, Additional mixing bowl: *Stainless steel*, Additional discs: *discs for vegetables*, Measuring cup: *Included*, Mincer: *Included with shortbread biscuits attachment*, Further attachments: *Blender*, Recipe book: *Low carb*
Study 3	Brand: *Fitbit*, GPS: *Integrated*, Heart rate monitor: *Integrated*, Sleep detection: *Yes*, Food tracker app: *Included*, Fitness coach app: *Not included*
Supplemental Study A	Hard disk capacity: *1TB*, Color: B*lack*, Number of controllers: *Two dual shock controllers*, Accessories: *Charging station*, Action-adventure games: *GTA V*, Role-playing games: *The Witcher 3*, Games for family & companionship: *FIFA 16*

We ordered the reservation prices from low to high values and computed the fraction of participants who would still buy for each possible price level (one € steps). Figure H2 shows the resulting (non-parametric) demand curves. Vertical dashed lines represent the median reservation price for each curve.

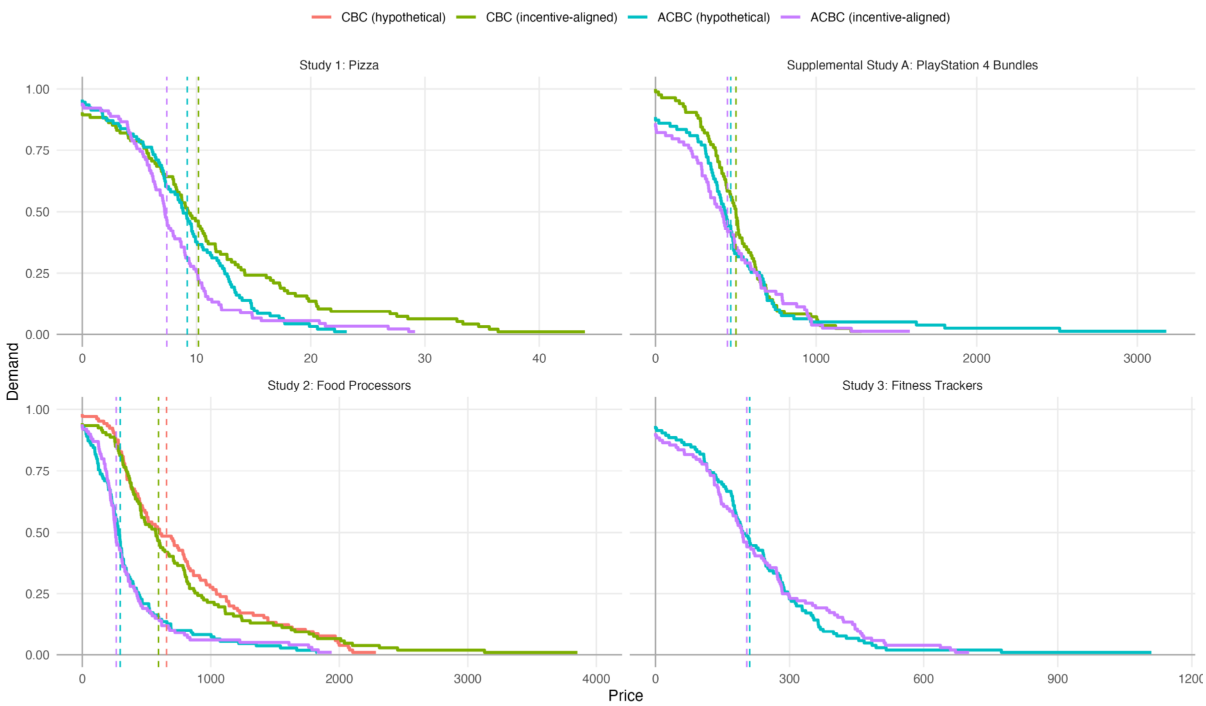


Fig. H2. Demand curves from reservation prices

We summarize the results as follows: (1) All demand curves have a similar shape (i.e., downward sloping sigmoid, with not all participants buying at price = 0 and a long tail with rather extreme reservation prices). (2) Adaptive designs and incentive alignment appear to lower (median) reservation prices (vertical dashed lines), but adaptive designs have a more pronounced effect compared to incentive alignment. (3) Conditions without incentive alignment have some participants with extremely high and unrealistic reservation prices (e.g., Study 2 > €2000). (4) Differences between the ACBC conditions are fairly small.

Generally, the order of conditions based on median reservation prices is consistently the same across studies: $\overline{rp}^{hCBC} > \overline{rp}^{iCBC} > \overline{rp}^{hACBC} > \overline{rp}^{iACBC}$. Higher reservation prices for CBC do not necessarily result from flatter curves as in Studies 1 and 2. In Supplementary Study A (PlayStation 4 bundles), the demand curve in the CBC condition is actually steeper compared to the ACBC conditions, but also, the fraction of participants accepting the product at a price of €0 is higher.

Table H3. Median reservation prices in € across studies and conditions

Condition	Fraction of Participants with $rp_i > 0$	Median of $rp_i > 0$	Median of $rp_i > 0$, averaged across all possible products
Study 1 (Pizza)			
iCBC	0.89	10.17	7.01
hACBC	0.95	9.19	5.43
iACBC	0.93	7.40	5.10
Supplemental Study A (PS4)			
iCBC	0.99	501.07	330.48
hACBC	0.87	468.58	266.02
iACBC	0.85	447.95	288.11
Study 2 (Food Processors)			
hCBC	0.97	655.71	370.31
iCBC	0.93	593.05	356.73
hACBC	0.93	295.17	192.97
iACBC	0.93	264.18	168.79
Study 3 (Fitness Trackers)			
hACBC	0.92	210.81	96.01
iACBC	0.89	204.22	97.89

Table H3 summarizes the aggregate results. While in all studies and conditions, the fractions of participants with positive reservation prices are quite high (85% - 99%), we observe managerial relevant differences in reservation prices. In particular, the results from the ACBC conditions are approximately 10% (Supplemental Study A) to 50% (Study 2) smaller compared to CBC conditions.

We also repeated the analysis using all possible combinations of attributes and their levels (cartesian product, see Gensler et al. (2012) for a similar approach). We calculated the median reservation price for each combination and averaged the results within the conditions. This is an elaborate step, as our choice designs are quite complex. For example, Study 2 (Food processors) allows for $3 \times 2 \times 4 \times 4 \times 2 \times 4 \times 6 \times 6 = 27{,}648$ combinations. For each combination, within a study and condition, we computed median reservation prices and then averaged the results across all combinations. The last column in Table H3 shows the results. The resulting reservation prices are lower but still reasonable, which intuitively makes sense, as we also include many "unpopular" products. The order of the results often remained the same (e.g., Study 1 and 2). Only in Study 3 and Supplemental Study A, the incentive-aligned ACBC conditions have slightly higher values compared to the hypothetical counterpart.

Lastly, we also repeated the analysis for the popular product without participants with extreme responses ("never" or "always" none-choosers) in the CBC conditions (see Web Appendix G). Figure H3 shows the results, with the same scaling for the x-axis to facilitate comparability.

We see that rather extreme reservation prices for the CBC conditions in Study 2 are gone. However, the shapes of the curves in Study 1 and Supplemental Study A are nearly identical. Note that pruning the data in the CBC conditions leads to demand curves that start at 100% for a price of €0. In Study 1 and Supplemental Study A, the resulting median reservation prices are slightly lower (Study 1: from €10.17 to €10.04 (-1.3%) and Supplemental Study A: from €501.07 to €483.10 (-3.6%)). In Study 2, the reductions are more substantial (hCBC: from €655.71 to €424.07 (-35.3%) and iCBC: from €593.05 to €391.43 (-34.0%)). Note that the aggregate CBC reservation prices are still larger than the ACBC results, and the order of the results is not affected. Hence, we conclude that pruning the CBC data leads to much more restricted distributions of reservation prices. This is in line with the motivation and the results of the procedure in Gensler et al. (2012). While the values are now somewhat more realistic in magnitude, they are still higher compared to the ACBC results. Here, no ad hoc pruning of the data is necessary, which seems to be an advantage. Lastly, the ACBC conditions still have some larger values for the reservation prices. We think that having a few consumers who are willing to pay a lot for a popular product is not unrealistic, and identifying these consumers can be crucial for personalized marketing (Allenby and Ginter 1995).

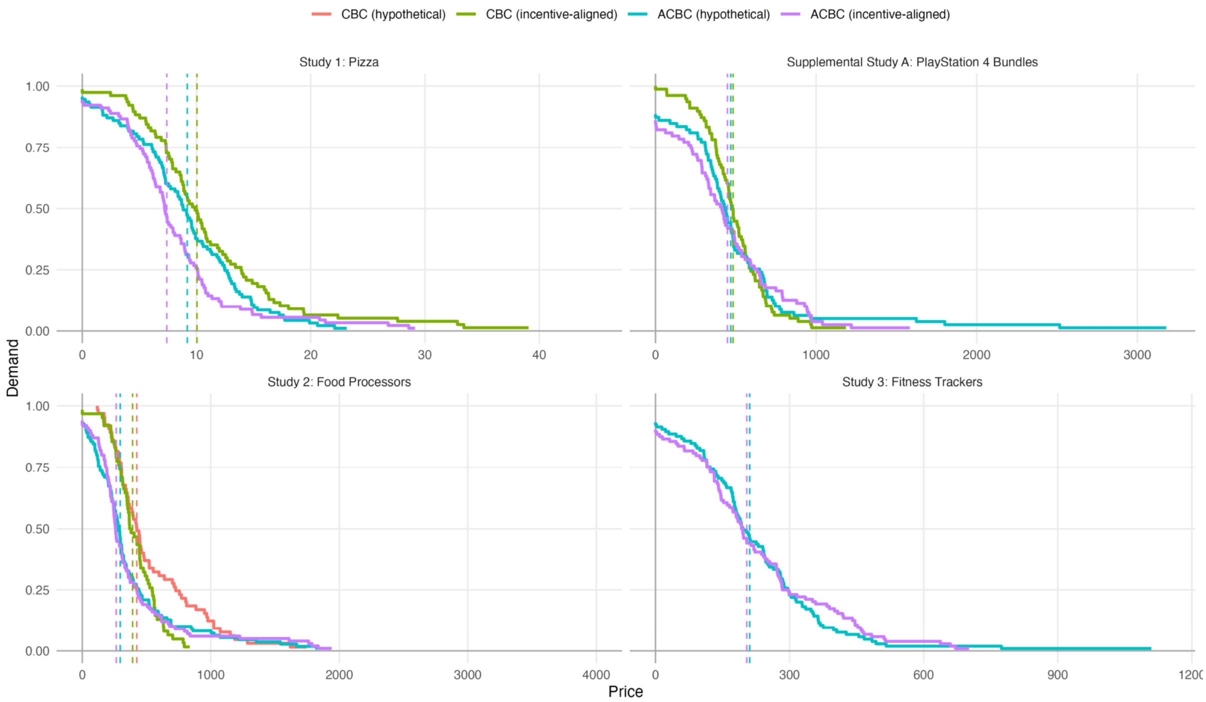


Fig. H3. Demand curves from reservation prices (without extreme responses in CBC)

Price elasticities

The analysis of the reservation prices has shown that the different conditions and, therefore, different design decisions can affect marketing implications. While we have already shown that differences in (the distribution of) price parameters exist (e.g., average and median values, amount of heterogeneity, fraction of wrong signs in the case of no sign constraint, see Web Appendix G), directly comparing them is difficult. However, it seems reasonable to expect that besides outside good-shares, the price sensitivity of participants can be a relevant driver of the differences in reservation prices. Therefore, we analyze the price elasticities of the different conditions in each study, as they are directly comparable and important quantities for managers, as they directly inform optimal pricing decisions.

We investigate two scenarios: (1) We compute price elasticities based on the "popular" products we already used to analyze the reservation prices. We use the average median reservation prices per study as price values at which we evaluate the elasticities. While this analysis is informative for understanding the price sensitivity in that particular situation, it does not account for secondary demand effects of prices (e.g., switching between products), as the outside good is the only competitor. 2) To address this, we additionally compute price elasticities for the HOT in each study (HOT 3 in Study (2), using the specific products and prices participants saw in these situations. Here we have 16, 6, 9, and 12 products in each choice set in Studies 1, 2, 3, and Supplemental Study A, respectively.

In both scenarios, we compute the (own) price elasticity for each participant:

$$\eta_{ij} = -\beta_i^p \cdot price_j \cdot \left(1 - p_{ij}\right), \qquad \text{(formula H2)}$$

where p_{ij} is the logit choice probability of participant i for alternative j, and then average the results across participants (within each condition and study). In scenario 2, we further aggregate the product elasticities to one market-level elasticity by using choice shares as weights (i.e., average choice probabilities).

The results in Table H4 are interesting. First, we see that elasticities are overall smaller (in absolute terms) in scenario 1 compared to scenario 2. While all values are reasonable for the product categories at hand (see, e.g., Bijmolt et al. 2005), the missing competition between products is the main driver for this result (through the smaller term $1 - p_{ij}$). Second, we do find for Study 2 that price elasticities are stronger in ACBC conditions compared to those without adaptive designs. In scenario 2, this difference is even more pronounced. However, in the other studies, we see mixed results regarding this design feature, i.e., in Supplemental Study A the elasticity of the incentive-aligned CBC condition is between that of the ACBC conditions, whereas in Study 1 (scenario 2), the CBC elasticity is even the highest in comparison (even though all elasticities are very similar). Third, we do not see a consistent picture across all studies when focusing on the main effect of incentive alignment. Stronger elasticities are only present in Study 2 (for CBC and ACBC conditions). On the other hand, in Study 3 and Supplemental Study A, price elasticities are slightly higher (in absolute terms) for the hypothetical ACBC conditions.

Table H4. Aggregate (own) price elasticities across studies and conditions

Study	**Scenario 1** (one popular product vs. outside good)	**Scenario 2** (HOTs)
Study 1 (Pizza)		
iCBC	-1.128	-2.079
hACBC	-1.437	-2.017
iACBC	-2.002	-2.001
Supplemental Study A (PS4)		
iCBC	-1.801	-4.659
hACBC	-2.043	-4.977
iACBC	-1.617	-4.545
Study 2 (Food Processor)		
hCBC	-0.661	-2.084
iCBC	-0.857	-2.222
hACBC	-1.053	-3.632
iACBC	-1.408	-3.948
Study 3 (Fitness Trackers)		
hACBC	-1.801	-3.266
iACBC	-1.617	-2.846

Notes: Scenario 1 = “popular” products from reservation price analysis.

In conclusion, we do not find a clear relationship between different design decisions and price elasticity. Therefore, it is misleading to attribute differences in reservation prices to price sensitivity alone. In Study 2, the very strong negative association between (absolute) price sensitivity and reservation prices might be specific to the particular product category. In the other studies, adaptive designs do not just trigger more careful comparisons of the (monetary) values of the products, but the specific valuation of particular features has a higher relevance for the results (see, e.g., the results for the WTP reported in Tables D17-D20).

To investigate this last argument further, we also looked at cross-price elasticity. Here we only analyze Study 3, as we have competing brands as alternatives and not just different variants/bundles of the sample product as in the other studies. To summarize these cross elasticities from the off-diagonal values of the full 9×9 matrix, we use the vulnerability and competitive clout measures of Kamakura and Russell (1989). Vulnerability represents the sum of the squared cross-price elasticities of an alternative considering the price change of the competitors. Competitive clout is the sum of the squared cross-price elasticities of all other alternatives regarding a price change of a particular alternative. Simply put, competitive clout measures how much impact one alternative has on others, whereas vulnerability reflects how much an alternative is affected by the competitor.

Figure H4 plots the two measures against each other for the choice scenario of the HOT for both conditions. The size of the circles reflects the aggregated choice share (i.e., power) of each alternative in the choice set. Dashed lines are choice share weighted averages.

The figure shows agreement for some alternatives across conditions, i.e., circles have a similar position, such as the ones of alternatives 3, 5, or 7. Nevertheless, there are also notable differences. While alternative 6 has similar values for vulnerability in both conditions, based on the ACBC with incentive alignment, competitive clout is much higher. For alternative 2, it is the other way round; very little competitive clout in ACBC with incentive alignment and much higher without. While both conditions agree in terms of vulnerability across all alternatives, the overall competitive clout is higher for the ACBC with incentive alignment.

This result also shows that the different conditions can affect marketing relevant insights beyond predictive validity. However, given that the ACBC with incentive alignment has the best prediction, one might conclude that the particular marketing insights from this condition are most relevant (i.e., closest to the truth). For practical purposes, design choices matter.

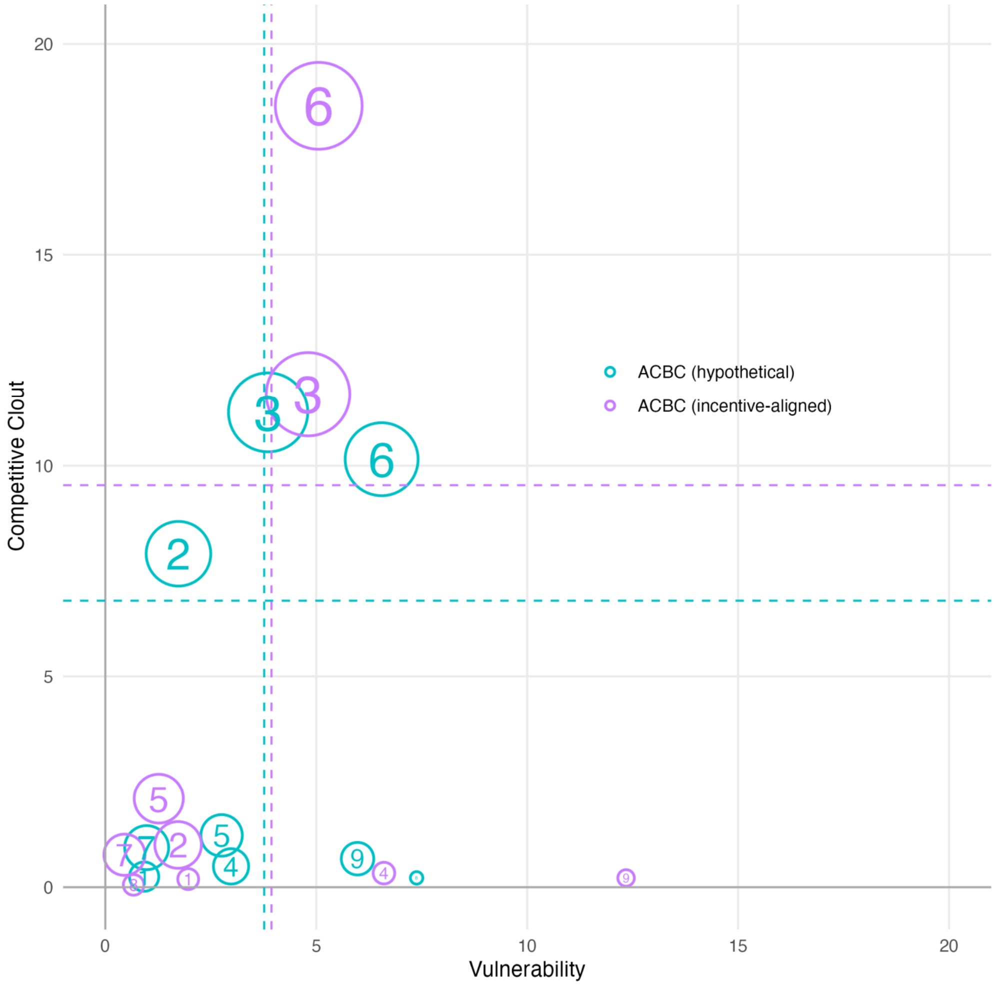


Fig. H4. Vulnerability and Competitive Clout in Study 3

The role of the fourth optional *Calibration* stage

To assess the effect of the 4th stage in the ACBC analyses, we first compare the posterior means of the none parameter after three and four stages. Note that this is reasonable as the other parameters are unaffected (see Web Appendix D for details). Therefore, the 4th stage also only affects predictions in free-choice HOTs (Supplemental Study A and Study 2 (HOT 3)). In both cases, the hit rates improved considerably (i.e., +12.02pp and +5.24pp, respectively); both increases are statistically significant with $p < 0.05$.

Table H5. Predictive validity of adaptive conjoint methods using 3 or 4 stages

	Hypothetical ACBC		Incentive-aligned ACBC	
HOT prediction (number of products, format)	**3 Stages**	**4 Stages**	**3 Stage**	**4 Stages**
Panel A: Supplemental Study A (12, free-choice)				
Hit rate in %	37.97	44.30 (+16.67%)	39.24	56.96 (+45.67%)
MHP in %	31.94	43.43 (35.97%)	38.46	53.33 (+38.66%)
Cohen’s kappa	0.23	0.20 (-13.04%)	0.25	0.42 (+68.00%)
MAE in %	4.24	1.65 (-61.08%)	3.98	1.52 (-61.81%)
Panel B: Study 2				
HOT 3 (6, free-choice)				
Hit rate in %	32.73	35.45 (+8.31%)	38.00	46.00 (+21.05%)
MHP in %	34.32	34.62 (+0.87%)	36.01	42.50 (+18.02%)
Cohen’s kappa	0.08	0.13 (+62.50%)	0.12	0.22 (+83.33%)
MAE in % (valid. sample)	5.88	5.58 (-5.10%)	6.50	6.40 (-1.54%)
HOT 6 (4, free-choice)				
MAE in % (valid. sample)	9.90	9.53 (-3.74%)	10.43	10.20 (-2.21%)

Notes: Green (Red) colored entries = Improvement (Deterioration) of the predictive validity metric when using 4 stages. Values in parenthesis are relative changes in %.

Table H5 reports results regarding the other measures for hypothetical and incentive-aligned ACBC conditions. Descriptively, the effect increases are (relatively) larger with incentive alignment for hit rates, MHP, and Cohen’s kappa. While this is an interesting result, note that the *Calibration* stage itself was not directly incentive-aligned. However, incentive alignment does indirectly affect the selection of products shown for *Calibration* (i.e., higher relevance due to more deliberation in the previous stages).

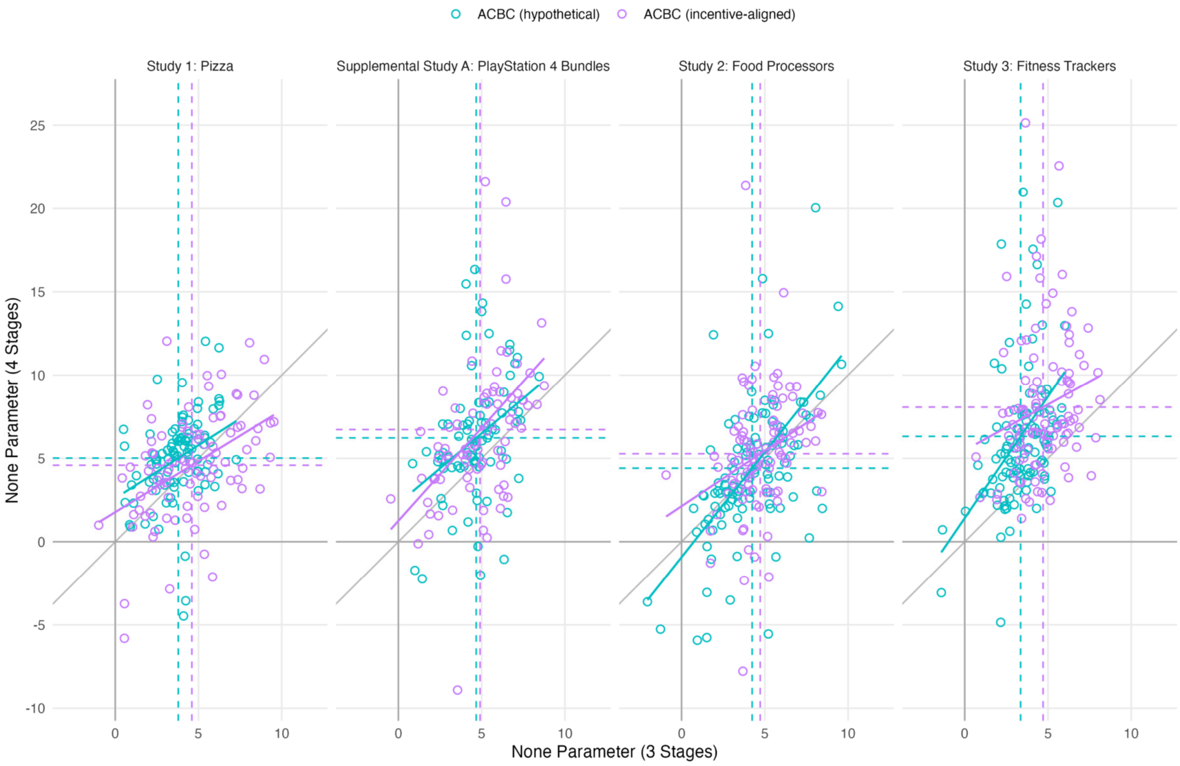
To aid our understanding of what specifically happens after the application of the *Calibration* stage, we next look at systematic changes in the individual none parameters. We analyze all four studies, as we routinely conducted the 4th stage in all studies. From the summary of the partworths in Tables D13-D16, we already learned that, except for one case, none parameters on average increased when implementing the *Calibration* stage. Hence, to some extent, higher outside-good shares could drive better predictive validity. Of course, we need to be cautious when comparing utility estimates directly, but this general tendency seems robust (note that CBC none parameters were even smaller). However, the additional information from the calibration tasks might also further affect the none parameters than "only" shifting them upwards. To investigate this, we regress the 4th stage none parameter on the 3rd stage none parameter and visually inspect their relationships.

Table H6. Results of none parameter regressions (stage 4 vs. stage 3)

Study	**Intercept (SE)**	**Slope (SE)**
Study 1		
hACBC	**2.606** (0.699)	**0.637** (0.173)
iACBC	**1.774** (0.645)	**0.611** (0.124)
Supplemental Study A		
hACBC	1.418 (1.100)	1.460 (0.303)
iACBC	**5.478** (1.297)	**0.551** (0.226)
Study 2		
hACBC	2.296 (1.237)	0.841 (0.251)
iACBC	1.254 (1.651)	1.116 (0.312)
Study 3		
hACBC	-0.929 (0.858)	1.259 (0.181)
iACBC	**2.130** (1.100)	0.666 (0.221)

Notes: Significant parameters with $p < 0.05$ are highlighted in **bold** and slopes are tested against 1.

Table H6 summarizes the results. If the 4th stage would only lead to a constant positive offset, we would expect a significant increase of the intercepts while having a slope of 1. On the other hand, if we expect the positive shifts to come from a relative increase, we would expect an intercept close to zero and a slope larger than 1. The results in Table H6 are mixed. While we see some (insignificant) intercepts close to zero and slopes larger than 1 (e.g., Supplemental Study A, hypothetical-ACBC), in most cases, we see positive intercepts and (positive) slopes smaller than 1 (e.g., Study 1). At least descriptively, this underlines that the 4th stage does more than a simple (relative) shift. Performing random effects single paper meta-analyses (McShane and Böckenholt 2017) for each incentive-aligned or hypothetical ACBC condition furnishes the conclusions regarding the slope effects quantitatively. For the hypothetical ACBC, we do not find a slope that is significantly different from 1 ($\beta = 1.020$, $z = 0.107$, $p = 0.915$), but Cochran's Q-test for heterogeneity (Hedges and Olkin 1985) is significant: $Q_{(3)} = 9.123$, $p = 0.028$. On the other hand, for incentive-aligned ACBC, the meta-analytic slope effect of 0.660 (i.e., -34%) is significantly smaller than 1 ($z = 3.568$, $p < 0.001$), but we find no indication of heterogeneity in this effect ($Q_{(3)} = 2.466$, $p = 0.482$).



Notes: In Studies 2, 3 and Supplemental Study A, we omitted one observation each with values for the none parameter after the 4th stage larger than 30.

Fig. H5. Estimated none parameters in ACBC conditions with 3 or 4 stages

Figure H5 speaks the same language. Dashed lines are average values of the none parameters, and the solid lines depict the fitted regression lines. In three out of four cases, we see for ACBC conditions with incentive alignment (pink), that slopes are less steep than the 45-degree line. At the same time, for studies 2 and 3, the average values of none parameters are higher after the 4th stage. Thus, the increase in none parameters is more (less) pronounced for participants with rather low (high) values (after three stages). Figure H5 also visualizes substantial variability, and many participants have also decreased none parameters after the 4th stage.

We interpret the result as indicating that, in many cases, the 4th stage does more than just shifting none parameters up (i.e., making free-choice predictions more conservative with regard to product choice vs. choice of the outside good (i.e., the option not to buy)). The additional information helps with "fine-tuning" the none parameters, with unequal effects regarding the direction and the magnitude of the shifts. We also implemented and tried a more sophisticated version of the 4th stage individual-level regressions. Specifically, we used a hierarchical linear model that shrinks the individual estimate towards the population means as this method shares information across participants. While we do not see "edge cases" anymore (e.g., slopes too small, intercepts close to zero, see Listing E4 in Web Appendix E for details), the results in terms of predictive validity were almost identical. This implies that although the hierarchical linear model should have better, more robust statistical properties, the information from the *Calibration* stage is "rich enough," so individual regressions (with ad hoc fixes for edge cases) work well in practice. Still, further research is needed to understand the 4th stage fully. One particular idea could be to *jointly* estimate the choice model for stages 1, 2, and 3, as well as the 4th *Calibration* stage, using a similar approach as Bacon and Lenk (2012) for combining choice and ratings data. This way, the information from the 4th stage would also affect population-level estimates (i.e., the upper model including correlation and uncertainty) and not just the posterior means of the participants.

Detailed results of the single paper meta-analyses

The single paper meta-analysis (SPM) enables analyzing the results of multiple studies within a single paper jointly (McShane and Böckenholt 2017). Complementing the core findings reported in the main article and the Web Appendix H, this subsection provides a more comprehensive presentation of the results in the form of forest plots and tests for heterogeneity.

Specifically, we summarize the full results for differences in MHP (Fig. H6 – H8), *BYO* times (Fig. H9), in study evaluation scores (Fig. 10), utility-scale (Fig. H11 – H14), and log-transformed times per choice tasks for each stage (Fig. H15 – H18) and overall due to incentive alignment (Fig. H19). These plots illustrate the effect estimates of the individual studies, along with their associated confidence intervals. The size of the square represents the weight that the corresponding study contributes to the SPM, i.e., the larger the square the larger the weight. The meta-analytical integrated effect of the considered studies is represented by the four-sided polygon at the bottom of the plot.

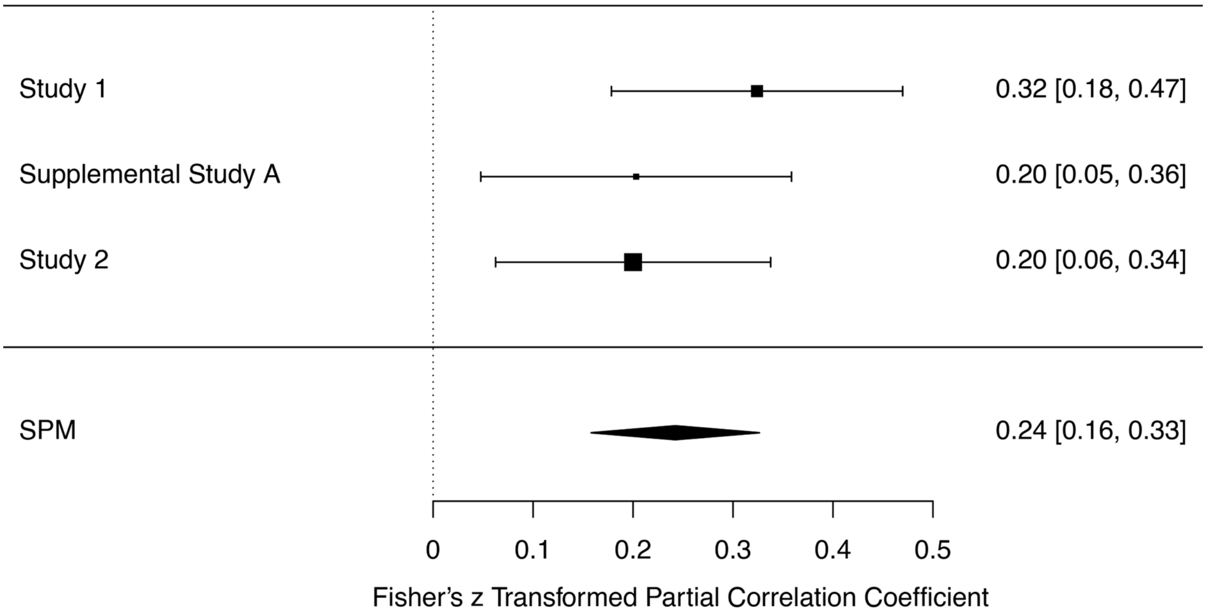


Fig. H6. Random effects SPM on the change in predictive validity as measured by the MHP (incentive-aligned ACBC vs. incentive-aligned CBC, total $n = 555$)

The forest plot in Fig. H6 visualizes the Fisher's z-transformed partial correlations as effect sizes between the implementation of adaptive designs in incentive-aligned conjoint analysis (i.e., incentive-aligned ACBC vs. incentive-aligned CBC) and the MHP. The individual studies, including Study 1, Supplemental Study A, and Study 2, show effect sizes of 0.32, 0.20, and 0.20, respectively, each statistically significant. The SPM demonstrates a significant meta-analytical overall effect size of 0.24 (95% CI = 0.16 to 0.33), suggesting a superior impact of the adaptive design on predictive validity ($r_{\text{z-transformed, partial}} = 0.242$, $z = 5.68$, $p < 0.001$). The non-significant Cochran's Q-test for heterogeneity ($Q(2) = 1.84$, $p = 0.184$) indicates the consistency of these results across studies.

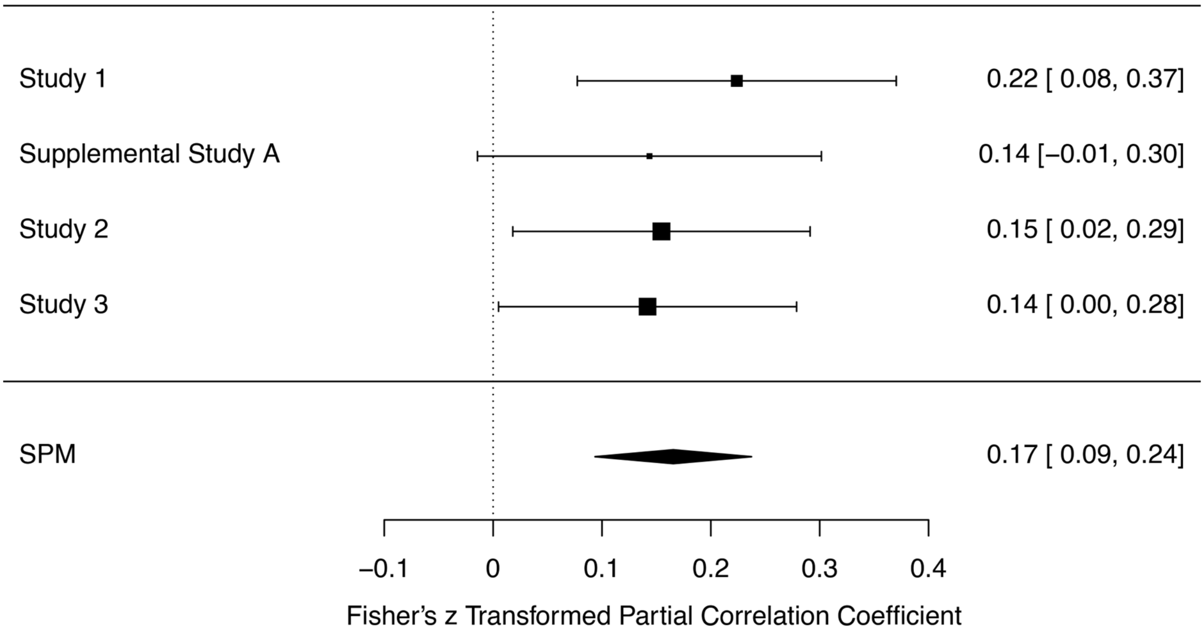


Fig. H7. Random effects SPM on the change in predictive validity as measured by the MHP (incentive-aligned ACBC vs. hypothetical ACBC, total $n = 760$)

Fig. H7 shows the equivalent SPM concerning the use of an incentive-aligned design over a hypothetical one in adaptive choice-based conjoint analysis (i.e., incentive-aligned ACBC vs. hypothetical ACBC) and the corresponding impact on MHP. The individual studies yield effect sizes in the span of 0.14 to 0.22. The meta-analytical effect size is statistically significant ($r_{\text{z-transformed, partial}} = 0.17$, $z = 4.54$, $p < 0.001$). The lack of significance in Cochran's Q-test for heterogeneity ($Q(3) = 0.83$, $p = 0.842$) implies uniformity of this effect across varying incentive mechanisms and product categories.

Fig. H8 continues the examination regarding a hypothetical ACBC vs. an incentive-aligned CBC, where effect sizes from individual studies fall between 0.03 and 0.10, all of which fail to achieve significance. The SPM effect is also not significant, with an effect size of 0.06 ($z = 1.51$, $p = 0.131$), suggesting a similar performance of the two methods. A non-significant Cochran's Q-test for heterogeneity ($Q(2) = 0.53$, $p = 0.769$) validates this consistency across the different studies.

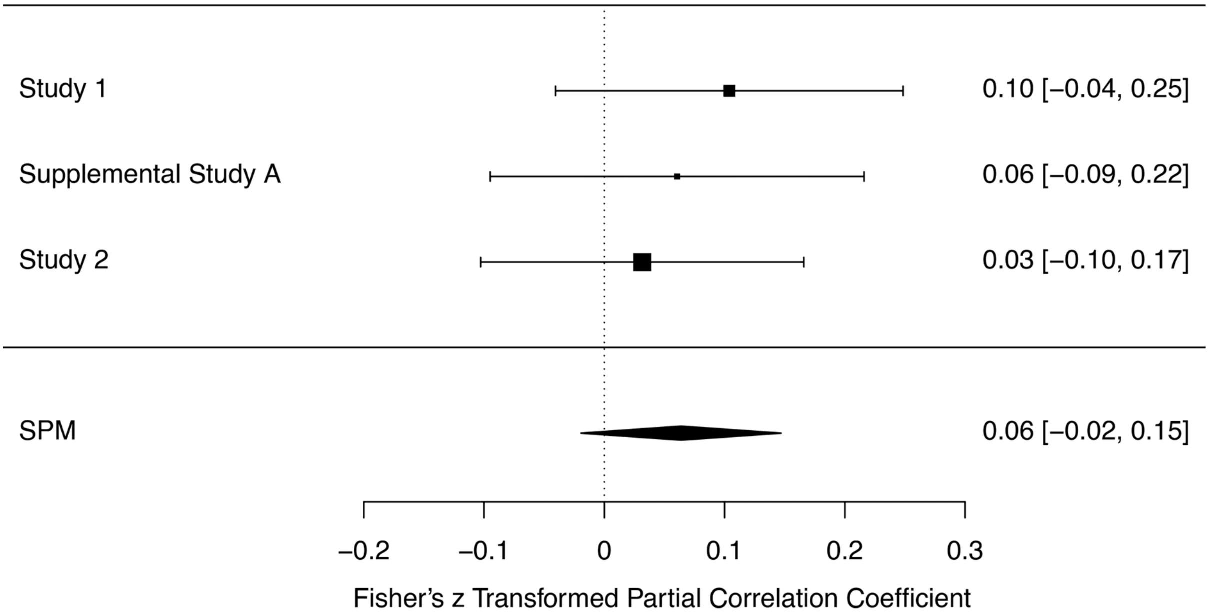


Fig. H8. Random effects SPM on the change in predictive validity as measured by the MHP (total n = 555)—of hACBC vs. iCBC

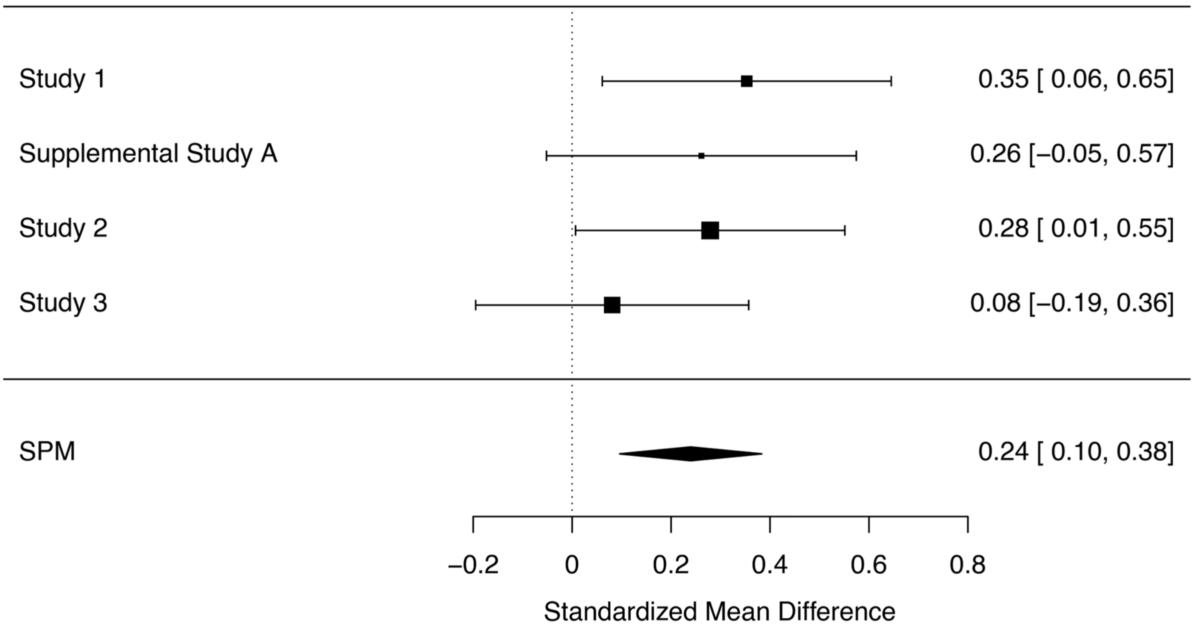


Fig. H9. SPM of log-transformed *BYO* time differences between iACBC vs. hACBC (total n = 753)

Fig. H9 illustrates the standardized mean differences in log-transformed *BYO* times between hypothetical and incentive-aligned ACBC designs. The effects of individual studies vary slightly, ranging from 0.08 to 0.35. Notably, two studies (Supplemental Study A and Study 3) have confidence intervals extending into the negative range. Still, overall, the meta-analytic effect is significant and positive (Hedges' g = 0.24, z = 3.27, p = 0.001). This suggests that, in general, incentive alignment leads to more deliberation in ACBC. A non-significant Cochran's Q-test ($Q_{(3)} = 1.95$, $p = 0.583$) further supports the homogeneity of this effect across different study contexts.

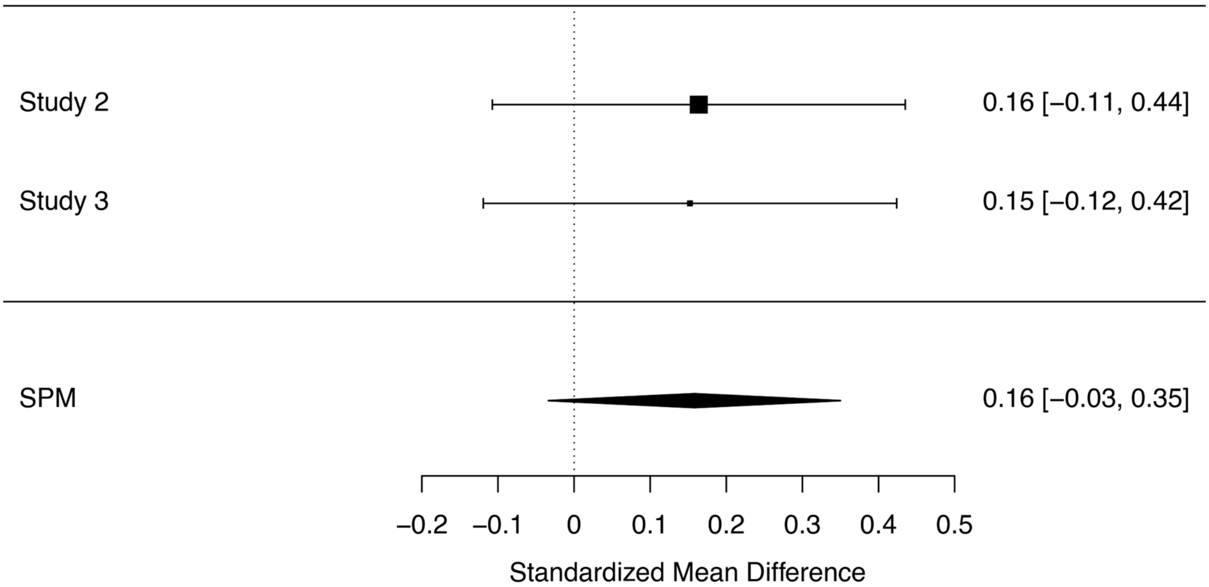
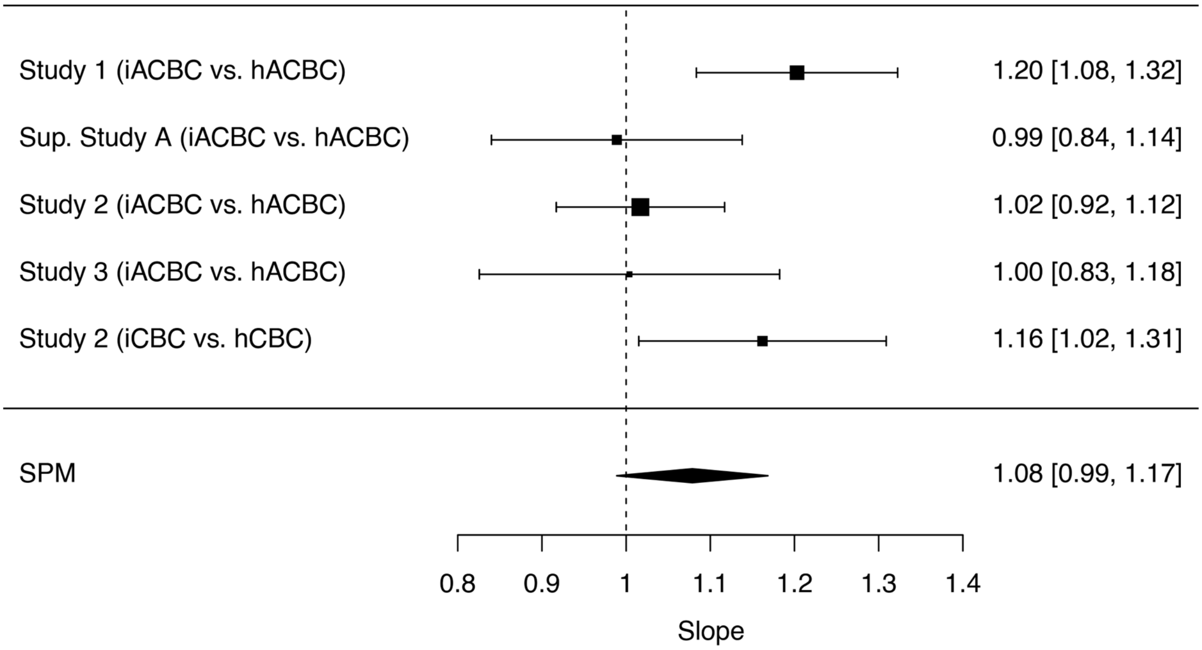


Fig. H10. SPM of the difference in the study evaluation scores between iACBC vs. hACBC (total $n = 419$)

As depicted in Fig. H10, the standardized mean differences in study evaluation scores between hypothetical and incentive-aligned ACBC variants in two studies were 0.16 and 0.15, respectively. The individual effects' confidence intervals cover the value zero, as well as the meta-analytic effect, which does not achieve statistical significance (Hedges' g = 0.16, $z = 1.61$, $p = 0.106$). The consistency of these results is affirmed by the low Cochran's Q-test value ($Q_{(1)} \approx 0$, $p = 0.953$).

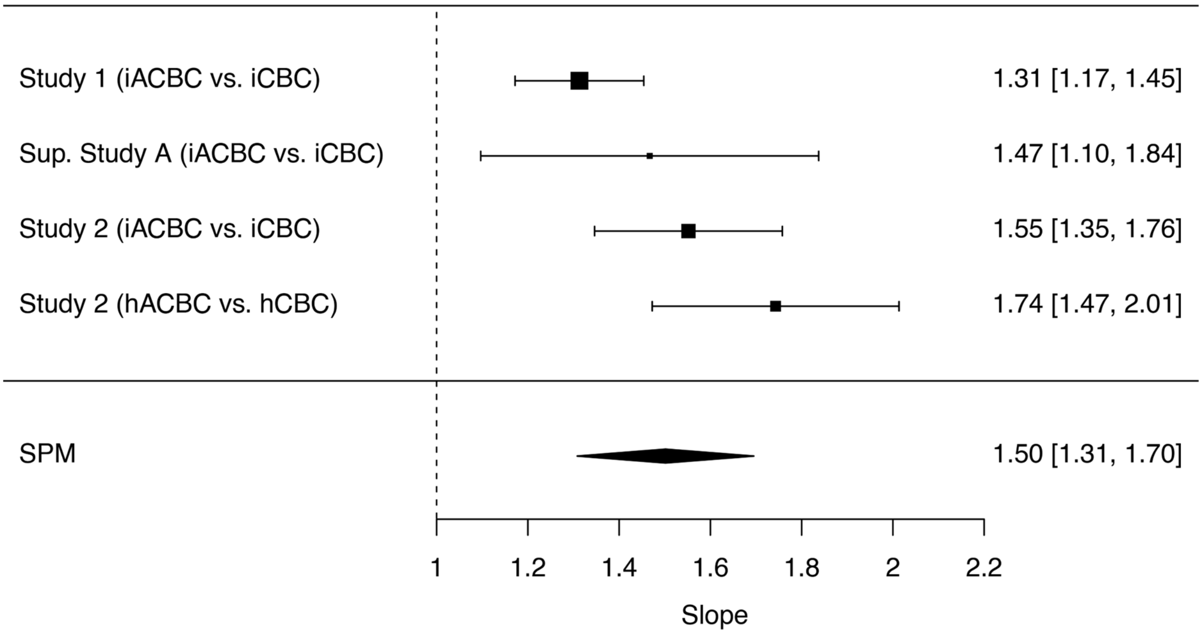


Notes: A slope of 1 implies no scale differences.

Fig. H11. SPM of the slopes in the regression of the mean partworths in incentive-aligned vs. hypothetical ACBC conditions to assess possible scale differences

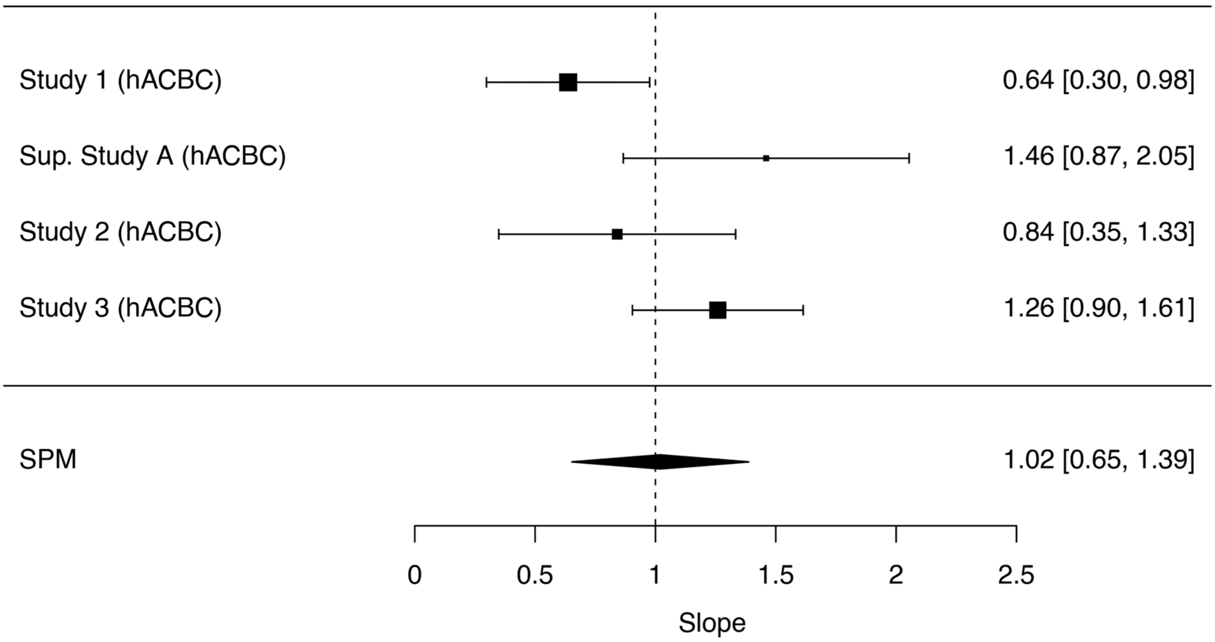
As displayed in Fig. H11, the slopes in the regression of the mean partworths (incentive alignment vs. hypothetical) range from 0.99 to 1.20 in the individual studies, with three studies showing non-significant results as their confidence interval crosses 1. The combined effect size is 1.08, which, however, does not statistically differ from a slope of 1 ($z = 1.72$, $p = 0.086$), suggesting no scale differences between the comparisons of the considered conditions. The Cochran's Q-test for heterogeneity is insignificant ($Q(4) \approx 8.916$, $p = 0.063$).

As shown in Fig. H12, the slopes in the regression of the mean partworths (ACBC vs. CBC) range from 1.31 to 1.74 in the individual studies, with no study crossing below 1 in their confidence intervals, thus suggesting significant differences. The cumulative effect size (SPM) is 1.50 and is statistically different from a slope of 1 ($z = 5.082$, $p < 0.001$), indicating significant scale increases across conditions with adaptive designs. However, the studies display significant heterogeneity, as shown by the Cochran's Q-test value of 9.076 ($p = 0.028$).



Notes: A slope of 1 implies no scale differences.

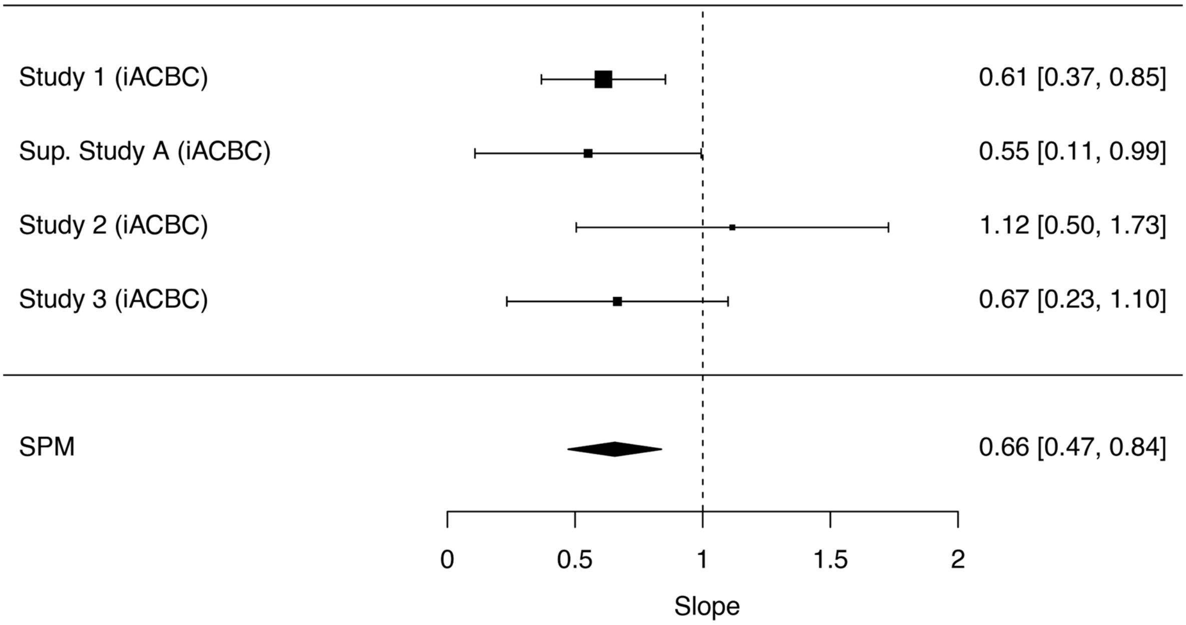
Fig. H12. SPM of the slopes in the regression of the mean partworths in ACBC vs. CBC conditions with incentive alignment to assess possible scale differences



Notes: A slope of 1 implies no scale differences.

Fig. H13. SPM of the slopes of the stage (stage 4 vs. stage 3) variable in the none parameter regression (hACBC)

Fig. H13 presents the none parameter regression slope analysis between stage 4 and stage 3 in hypothetical ACBC conditions. Study 1 has the smallest slope value of 0.64, while the largest one, 1.46, is seen in Supplemental Study A. The SPM exhibits a slope of 1.02 and a CI that crosses 1 (0.65 to 1.39), implying non-significance ($z = 0.107$, $p = 0.915$). The Cochran's Q-test points out a significant heterogeneity across the hACBC conditions ($Q_{(3)} \approx 9.123$, $p = 0.028$).



Notes: A slope of 1 implies no scale differences.

Fig. H14. SPM of the slopes of the stage (stage 4 vs. stage 3) variable in the none parameter regression (iACBC)

Fig. H14 compares the none parameter regression slopes for incentive-aligned ACBC conditions with 4 stages vs. 3 stages. Of the four studies analyzed, three exhibit mean slopes under 1, suggesting a decrease in scale in the transition from 3-stage to 4-stage incentive-aligned ACBC. The SPM demonstrates a significant slope of 0.66 with a CI of 0.47 to 0.84 ($z = 3.568$, $p < 0.001$), suggesting a significant decrease in scale. Although Study 2 shows a slight increase, the Cochran's Q-test provides no evidence of significant heterogeneity among the studies ($Q_{(3)} \approx 2.466$, $p = 0.482$).

Fig. H15 presents the comparison of the effects of incentive alignment on the log-transformed times per choice task in the *BYO* stage. We can interpret the difference approximately as percentage changes. Of the four studies analyzed, three exhibit longer times per task suggesting more deliberation in ACBC conditions with incentive alignment. The SPM shows a significant total effect of 0.092 with a CI of 0.014 to 0.170 ($z = 2.323$, $p = 0.020$), suggesting a significant increase in deliberation. The Cochran's Q-test provides evidence of significant heterogeneity among the studies ($Q(3) \approx 7.845$, $p = 0.049$).

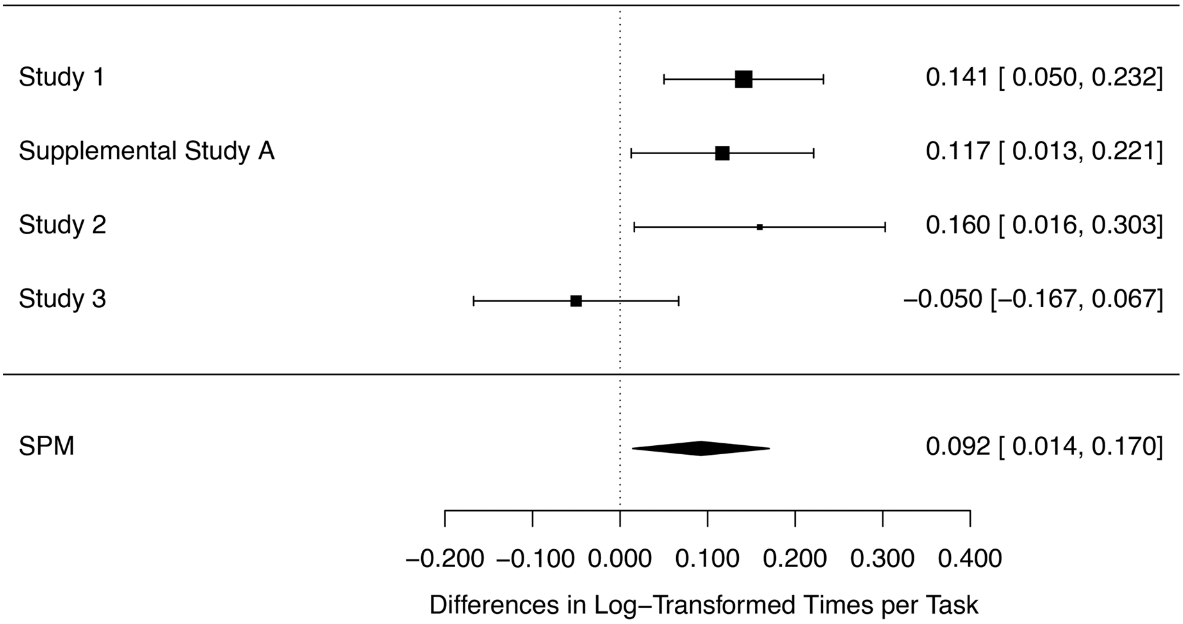


Fig. H15. SPM of differences in log-transformed times per task between iACBC vs. hACBC in the *BYO* stage

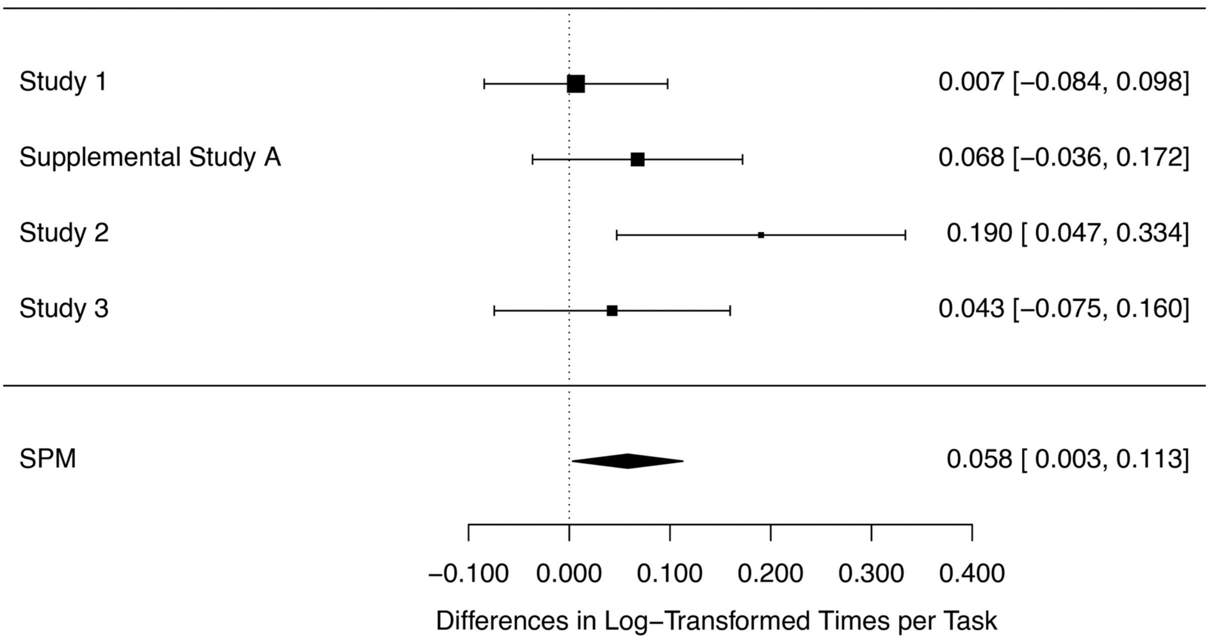


Fig. H16. SPM of differences in log-transformed times per task between iACBC vs. hACBC in the *Screening* stage

Fig. H16 summarizes the comparison of the effects of incentive alignment on the log-transformed times per choice task in the *Screening* stage. All studies analyzed have longer times per task, suggesting more deliberation in ACBC conditions with incentive alignment. The SPM shows a significant total effect of 0.058 with a CI of 0.003 to 0.113 ($z = 2.0789$, $p = 0.038$), suggesting a significant increase in deliberation. According to the Cochran's Q-test, there is no evidence of significant heterogeneity among the studies ($Q(3) \approx 4.606$, $p = 0.203$).

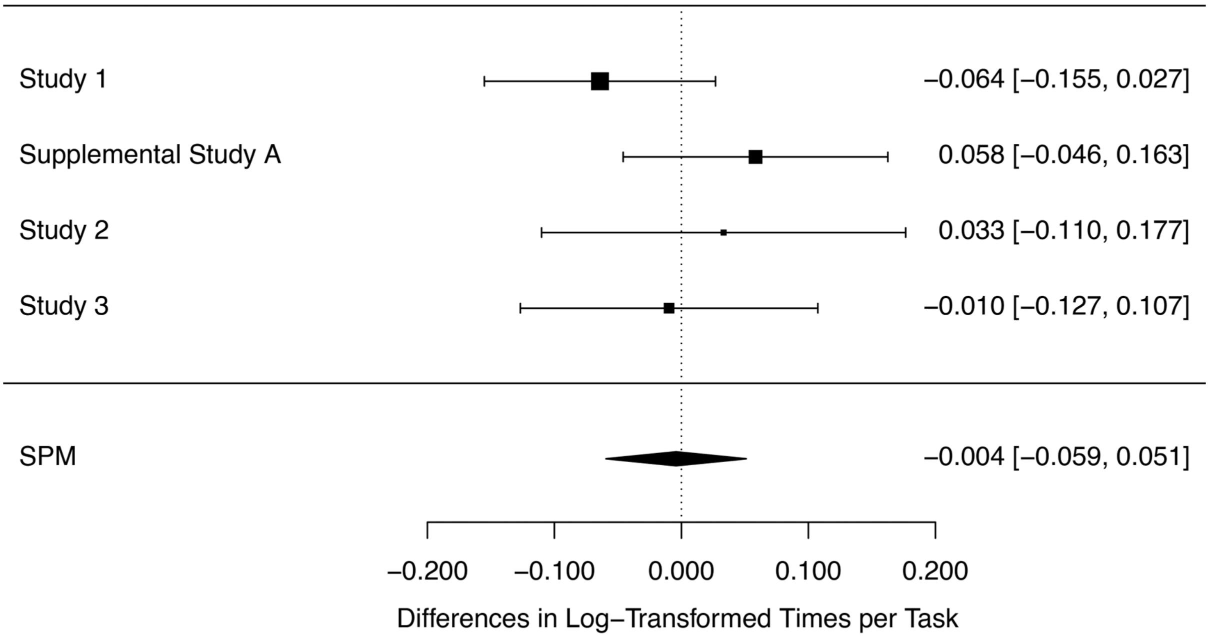


Fig. H17. SPM of differences in log-transformed times per task between iACBC vs. hACBC in the *Tournament* stage

Fig. H17 compares the effects of incentive alignment on the log-transformed times per choice task in the *Tournament* stage. Two of the four studies have longer times per task and for the other two, the times per task are shorter. Hence, we find no clear result regarding deliberation in ACBC conditions with incentive alignment for the *Tournament* stage. The SPM shows an insignificant total effect of -0.004 with a CI of -0.059 to 0.051 ($z = -0.149$, $p = 0.882$). According to the Cochran's Q-test there is no evidence of significant heterogeneity among the studies ($Q(3) \approx 3.322$, $p = 0.345$).

Fig. H18 depicts the comparison of the effects of incentive alignment on the log-transformed times per choice task in the *Calibration* stage. Three of the four studies have longer times per task. The SPM shows an insignificant total effect of 0.033 with a CI of -0.021 to 0.088 ($z = 1.191$, $p = 0.234$), providing only directional support for an increase in deliberation. Based on the Cochran's Q-test there is no evidence of significant heterogeneity among the studies ($Q(3) \approx 3.667$, $p = 0.300$).

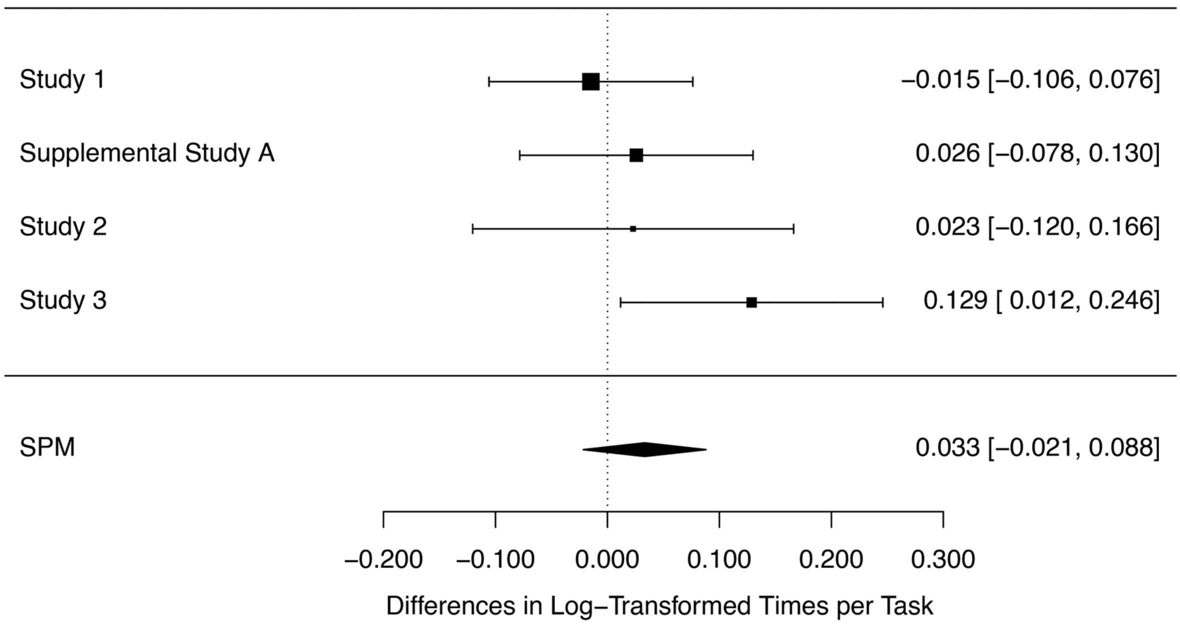


Fig. H18. SPM of differences in log-transformed times per task between iACBC vs. hACBC in the *Calibration* stage

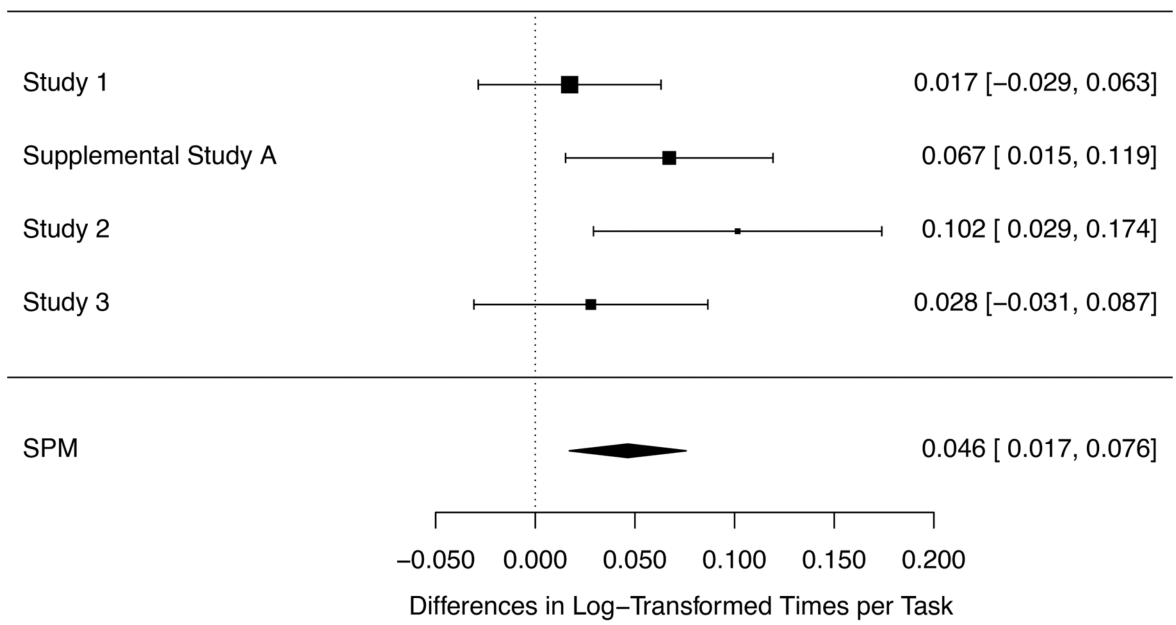


Fig. H19. SPM of log-transformed times per task differences between iACBC vs. hACBC across all stages

Fig. H19 compares the effects of incentive alignment on the log-transformed times per choice task across all stages. All four studies have longer times per choice task in ACBC conditions with incentive alignment. The SPM shows a significant effect of 0.046 with a CI of 0.017 to 0.076 (z = 3.105, p = 0.002), suggesting a significant increase in deliberation of about 5%. The Cochran's Q-test confirms heterogeneity among the studies ($Q_{(3)} \approx 4.780$, p = 0.189).

Reanalysis of Hauser et al. (2019) data

Hauser et al. (2019) investigated the impact of image realism and incentive alignment on scale (see also *"Scale differences across conditions"* in this Web Appendix). To this end, the authors conducted an experiment with four conditions: (1) "text only, no incentive alignment," (2) "text only, with incentive alignment," (3) "realistic images, no incentive alignment," and (4) "realistic images, with incentive alignment." Despite having four conditions, hit rates were only reported for both extremes of the experimental conditions, i.e., first and fourth condition. The missing hit rates for (2) and (3) may be attributable to the study's main focus, which was the effect of different CBC designs on scales, rather than hit rates. Therefore, we reanalyzed the data to obtain the missing hit rates using the replication files provided by the authors as supplementary material. Furthermore, we performed a series of two-proportions Z-tests. The results are presented below.

(1) Text only, no incentive alignment (hit rate = 0.244, n = 275)
- vs Text only, with incentive alignment: z = -2.11, p = 0.035
- vs Realistic images, no incentive alignment: z = -2.88, p = 0.004
- vs Realistic images, with incentive alignment: z = -3.65, p < 0.001

(2) Text only, with incentive alignment (hit rate = 0.327, n = 251)
- vs Realistic images, no incentive alignment: z = -0.76, p = 0.449
- vs Realistic images, with incentive alignment: z = -1.48, p = 0.139

(3) Realistic images, no incentive alignment (hit rate = 0.360, n = 248)
- vs Realistic images, with incentive alignment: z = -0.71, p = 0.481

(4) Realistic images, with incentive alignment (hit rate = 0.392, n = 270)

The results show that only the paired comparisons of hit rates involving the “text only, no incentive alignment” condition yield statistically significant differences. In contrast, all other comparisons did not demonstrate statistical significance. Thus, the hit rate for “realistic images, with incentive alignment” (0.392) is statistically not different from both “text only, with incentive alignment” (0.327) and “realistic images, no incentive alignment” (0.360). This suggests that the combination of incentive alignment and the use of realistic images does not necessarily enhance predictive accuracy, as indicated by the hit rates, compared to using each approach independently.

Analysis of the superior performance of ACBC in free-choice HOTs

In this section, we analyze why ACBC outperforms CBC in particular in free-choice HOTs. Specifically, we analyze whether this is due to better predicting product choices in general (i.e., secondary demand) or purchase incidence (i.e., primary demand).[12] In two of our studies, we included free-choice HOTs (Supplemental Study A: PlayStation 4 Bundles and Study 2: Food Processors (HOT 3)). For these HOTs, we split the results for the mean hit probabilities depending on whether the choice of the participants (i.e., the hit) was for a product or the outside option (i.e., choosing not to buy any product).

Fig. H20 summarizes the results: (1) All variants of conjoint analysis demonstrate superior predictive validity regarding outside good choices compared to product choices. (2) Incentive alignment enhances the ability to predict product choices across all conjoint analysis variants accurately. (3) Notably, the introduction of incentive alignment clearly boosts the predictive validity for outside good choices. This finding aligns with the common view that incentive alignment reduces hypothetical bias (e.g., Miller et al. 2011). Surprisingly, previous studies on incentive-aligned CBC have largely overlooked this specific aspect. (4) Study 2 (HOT 3) further reveals that the incentive alignment-induced relative improvement in predictive validity for

[12] We thank one reviewer for suggesting this analysis.

outside good choices is much more pronounced for CBC (+62%) than for ACBC (+19%). (5) Even without incentive alignment, ACBC outperforms CBC in terms of predictive validity for outside good choices. Yet, this is not always observed when predicting inside good choices.

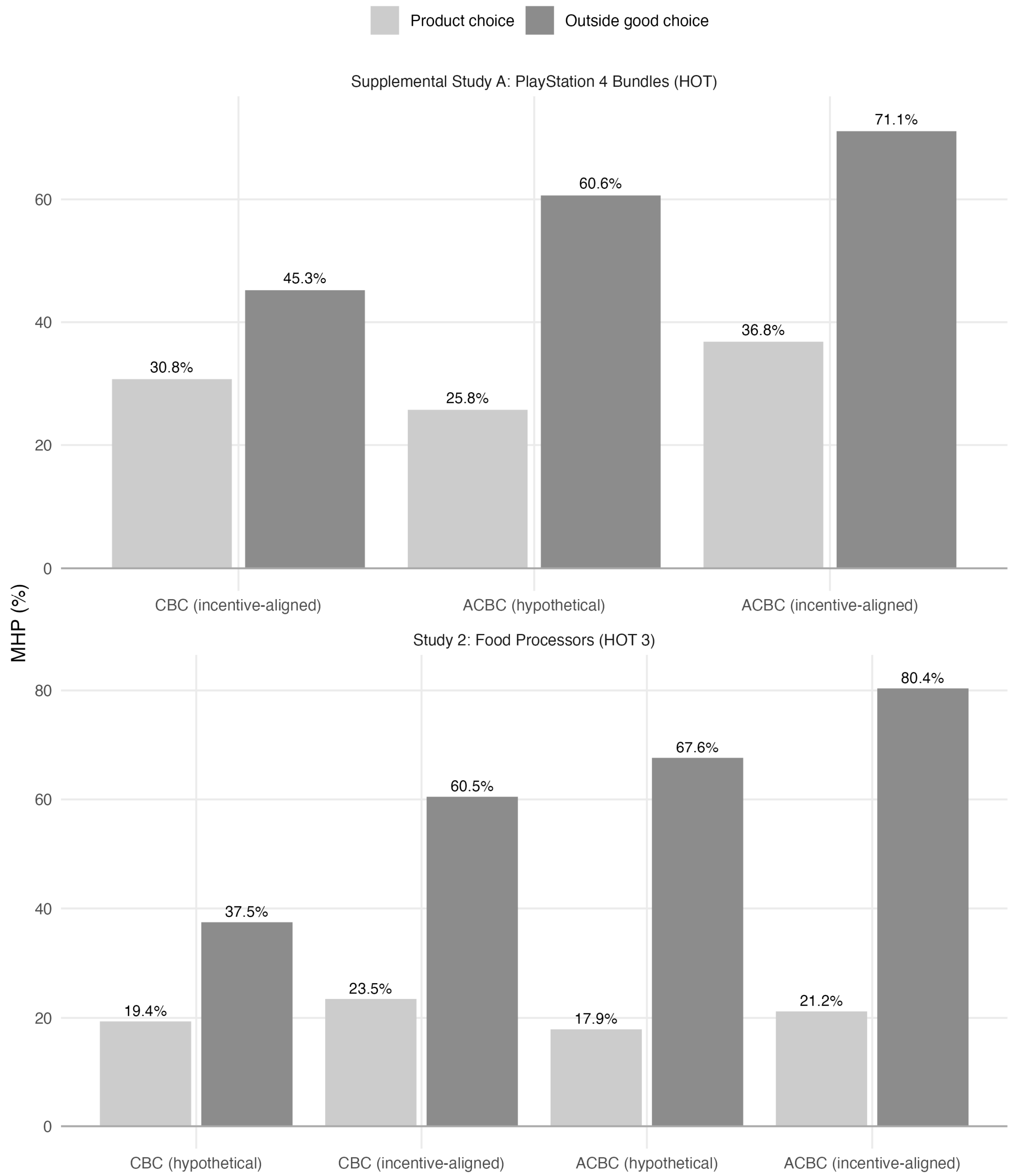


Fig. H20. Predictive validity in free-choice HOTs for product choice and purchase incidence

Our descriptive analysis of the MHP shows that the application of ACBC, as opposed to CBC, appears to yield larger gains in predictive validity from an increased capability to accurately predict outside good choices, rather than from a better prediction of product choices.

However, our analysis of other forced-choice HOTs reveals that ACBC not only has superior predictive validity when an outside good is included. The superiority is further amplified in HOTs that provide the option not to buy.

Comparison of survey times for all stages in ACBC studies

We compare the ACBC conditions with and without incentive alignment of all studies with regard to the time and the number of tasks needed for completing the different stages. This comparison sheds further light on how incentive alignment affects the choice behavior of participants in ACBC studies.[13]

Table H7 reports the mean (median) times for each ACBC condition (incentive alignment, hypothetical), stage (*BYO*, *Screening*, *Tournament*, *Calibration*), and study (Study 1: Pizza, Suppl. Study A: PlayStation 4 Bundles, Study 2: Food Processors, Study 3: Fitness Trackers).

Table H7. Survey times in seconds (mean (median)) in each study, condition, and stage

		Stage				
Study	**Condition**	**BYO**	**Screening**	**Tournament**	**Calibration**	**Total**
Study 1:	hACBC	81.6 (79.0)	357.5 (349.0)	167.4 (148.0)	50.9 (47.0)	657.37 (635.0)
	iACBC	96.6 (85.0)	365.7 (347.0)	132.0 (120.5)	49.7 (45.0)	643.97 (608.5)
Suppl. Study A:	hACBC	89.5 (85.0)	324.2 (292.0)	140.1 (128.0)	61.5 (57.0)	615.32 (548.0)
	iACBC	109.4 (88.0)	352.4 (307.0)	148.0 (136.0)	62.7 (59.0)	672.49 (575.0)
Study 2:	hACBC	167.5 (129.0)	509.7 (434.0)	216.5 (161.0)	77.6 (67.5)	971.31 (859.5)
	iACBC	189.9 (160.5)	654.2 (467.5)	224.0 (194.5)	79.2 (69.0)	1147.30 (1013.5)
Study 3:	hACBC	269.1 (190.0)	254.1 (208.0)	144.8 (98.0)	74.5 (54.0)	742.43 (573.0)
	iACBC	219.5 (195.0)	267.0 (231.0)	118.6 (99.5)	75.5 (63.5)	680.57 (617.0)

The table shows that in most cases across studies and stages, the avg. times for the conditions with incentive alignment are longer (highlighted in green). This is consistent with the argument that incentive alignment motivates more task deliberation. In some instances (e.g., *BYO* stage in the Fitness Trackers study), the hypothetical condition took longer. However, as the distribution for the times is right-skewed, the avg. value can be affected by a few extreme values. Indeed, most median values are smaller than the mean values, and provide more robust information. Still, even when looking at the median values, the *Tournament* and *Calibration*

[13] We thank a reviewer for suggesting a holistic analysis of the times in the ACBC conditions.

stage times are shorter in the Pizza study with incentive alignment. In general, the total times (last column) show the same pattern. Comparing the total times with the avg. (median) times for the CBC conditions, we find that the corresponding ACBC conditions are about 2 to 2.5 times longer (Study 1, iCBC: 275.74 (263); Suppl. Study A, iCBC: 357.44 (279); Study 2, hCBC: 435.73 (324); Study 2, iCBC: 566.62 (379)).

It is important to note that in ACBC studies, the design of the later stages (i.e., what and how to ask) is affected by the participants' answers in earlier stages. This is the reason why we only look at the difference in *BYO* times in the main paper, as the tasks in this stage do not differ across participants (or conditions) within a study (i.e., the number of non-price attributes). To analyze this point, we summarize in Table H8 the avg. number of choice tasks per study, condition, and stage.

Table H8. Average number of choice tasks in each study, condition, and stage

		Stage			
Study	**Condition**	**BYO**	**Screening**	**Tournament**	**Calibration**
Study 1:	hACBC	7.0	13.0	8.3	6.0
	iACBC	7.0	13.2	7.1	6.0
Suppl. Study A:	hACBC	7.0	10.5	6.9	6.0
	iACBC	7.0	10.6	6.9	6.0
Study 2:	hACBC	8.0	12.2	7.6	6.0
	iACBC	8.0	12.0	8.3	6.0
Study 3:	hACBC	6.0	9.9	5.6	6.0
	iACBC	6.0	10.2	5.7	6.0

The table shows that, except for the Food Processor study, the avg. number of tasks for the *Screening* stage is higher in conditions with incentive alignment. More tasks are asked if a participant is more critical regarding specific attributes (so-called "must have" and "unacceptable" attributes) based on the answers in previous tasks. We interpret this result as indicating more deliberation and critical engagement with the task. However, this could also explain why times, in general, are longer with incentive alignment in the *Screening* stage. Also notable are the results for the *Tournament* stage in the Pizza study. In the condition with incentive alignment, this stage has more than one task fewer in the condition with incentive alignment. Fewer tasks in the *Tournament* stage are an indication that fewer concepts are left at this stage that can be shown to the participant without restrictions. Again, this is in line with more deliberation in the previous stages. This can further explain, why the avg. (and median) times for the *Tournament* stage are shorter with incentive alignment in the Pizza study (see Table H7). Interestingly, the direction of the effect is reversed in the Food Processors Study, i.e., the avg. number of tasks in the *Tournament* stage is smaller in the hypothetical condition, but this is due to skewed distribution. Looking at the median number of tasks, there is no difference between conditions anymore for this stage and study. On the other hand, for the Pizza study, the median number of tasks in the *Tournament* stage is two tasks lower (7 vs. 9) in the incentive alignment condition. The number of tasks in the *Calibration* stage does not seem to be affected by the condition across studies.

The differences in the number of tasks across conditions require a more refined analysis. Therefore, we summarize the times per task in each stage and study. To account for the skewness of this measure, we take the (natural) logarithm before calculating avg. values.

Table H9. Average log times per choice tasks in each study, condition, and stage

		Stage			
Study	**Condition**	**BYO**	**Screening**	**Tournament**	**Calibration**
Study 1:	hACBC	2.4	3.3	3.0	2.1
	iACBC	2.5	3.3	3.0	2.1
Suppl. Study A:	hACBC	2.5	3.4	2.9	2.3
	iACBC	2.6	3.4	3.0	2.3
Study 2:	hACBC	2,8	3.6	3.1	2.4
	iACBC	3.0	3.8	3.2	2.5
Study 3:	hACBC	3.6	3.1	2.9	2.2
	iACBC	3.5	3.2	2.9	2.4

The values in Table H9 are quite similar, but we can observe a specific pattern emerging: After controlling for the different number of choice tasks, most stages with incentive alignment conditions have longer processing times, potentially due to increased deliberation. In some cases, the times are equally long. The only exception is the *BYO* stage for the Fitness Tracker, but we have already discussed that some extreme values in the hypothetical condition adversely affect the results. Furthermore, in the PlayStation 4 Bundles study, only the *BYO* stage appears to be affected.

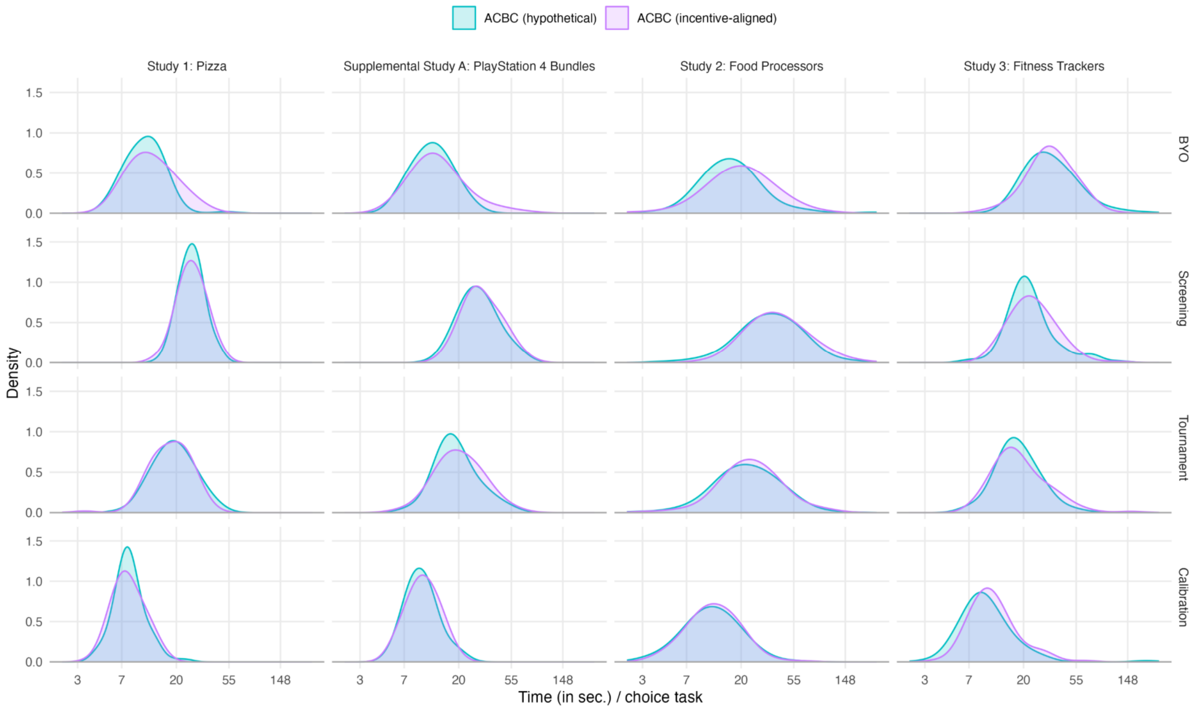


Fig. H21. Distribution of log times per choice tasks in each study, condition, and stage

Fig. H21 shows the full distributions of the results, where the x-axis (times in sec. per choice task) is scaled in logs. Generally, the resulting distributions look relatively symmetric and the highlighted mean differences from Table H8 are visible as shifts in the whole distributions (see, e.g., the *BYO* stage for Food Processors), even though most of the distributions are quite similar. We have already tested the effect of the conditions on the *BYO* times in the main paper (see also Web Appendix "Detailed results of the single paper meta-analyses" for the SPM results) and found a significant positive effect of incentive alignment. We repeat this analysis for the logarithm of the times per task to get a complete picture.[14] For each, we employ a mixed-effect model with random intercepts for stages and participants (to account for repeated measurement). The results show positive slope parameters for the

[14] Note that we also ran all analyses using the inverse normal transformation (instead of taking logs) to deal with the skewness of the dependent variable, and the results were very similar.

incentive alignment conditions in all studies, indicating an increase in time per task between 2% and 10%, but two are significant (Study 1: Pizza: $\beta = 0.02$, $t_{(645)} = 0.74$, $p_{\text{(one-tailed)}} = 0.\ 231$; Supp. Study: PlayStation 4 Bundles: $\beta = 0.07$, $t_{(549)} = 2.53$, $p_{\text{(one-tailed)}} = 0.006$; Study 2: Food Processors: $\beta = 0.10$, $t_{(749)} = 2.75$, $p_{\text{(one-tailed)}} = 0.003$; Study 3: Fitness Trackers: $\beta = 0.03$, $t_{(730)} = 0.93$, $p_{\text{(one-tailed)}} = 0.176$).

Based on the analysis above, it is evident that most of the positive effect of incentive alignment on task completion times stems from the *BYO* stage. By aggregating the results of the studies using SPMs, we can estimate the overall effect of the incentive alignment condition across all stages (refer to the Web Appendix *"Detailed results of the single paper meta-analyses"* for complete findings). These analyses reveal positive effects ranging from 3.3% in the *Calibration* stage to over 9.2% in the *BYO* stage. Only the *Tournament* stage exhibits a minimal negative effect of -0.4%, which is also statistically insignificant. Notably, the effects observed in the *BYO* and *Screening* stages are statistically significant (*BYO*: $\beta = 0.092$, $z = 2.323$, $p_{\text{(one-tailed)}} = 0.010$, *Screening*: $\beta = 0.058$, $z = 2.079$, $p_{\text{(one-tailed)}} = 0.019$; *Tournament*: $\beta = -0.004$, $z = -0.147$, $p_{\text{(one-tailed)}} = 1$; *Calibration*: $\beta = 0.033$, $z = 1.191$, $p_{\text{(one-tailed)}} = 0.117$). When we conduct the SPM without accounting for the stage-specific effect of the condition, we obtain a positive and significant effect of 4.6% ($\beta = 0.046$, $z = 3.105$, $p_{\text{(one-tailed)}} = 0.001$), providing robust evidence for increased deliberation through incentive alignment in ACBC studies.

In conclusion, the additional results support a positive effect of incentive alignment on the times per choice task in the different stages. The analysis also highlights that it is important to account for the endogenous nature of ACBC designs. A direct comparison of the total times is not always appropriate as the stages in ACBC studies can have different numbers of tasks depending on the answers of participants in "earlier" stages. After adjusting for this, we found significantly positive effects of incentive alignment in the *BYO* and the *Screening* stage. The effect for the other stages is positive, but not significant in all studies. The effect, however, is significant if we also pool over all stages. This furnishes further support for incentive alignment, leading to more deliberation in ACBC studies.

The effect of the amount of information used in estimation on the predictive validity

Next, we compare the effect of varying the number of choice sets in CBC conditions ($t = 2, \ldots, T$, where T is the total number of choice tasks in the study), and the number of stages in ACBC conditions ($s = 1, \ldots, 4$, i.e., *BYO*, *Screening*, *Tournament*, and *Calibration*) during the estimation on the gains in predictive validity (i.e., mean hit probability).[15] In the main paper and earlier in Web Appendix H, we have already analyzed the effect using the optional *Calibration* stage in free-choice HOTs. Here, we extend this analysis by further dropping the *Tournament* stage or even the Screening stage. That means we are either using only the first stage (*BYO*), the first and second stages (*BYO* and *Screening*), the first three stages (*BYO*, *Screening*, and *Tournament*), or all four stages (*BYO*, *Screening*, *Tournament*, and *Calibration*). Note that it would not make sense to keep any of the later stages while dropping an earlier stage because of the adaptive design, i.e., the choice tasks in later stages depend on the answers of earlier stages. A solution would be to set up experiments that test all kinds of permutations of ACBC stages as a between-subjects factor. In Web Appendix A, we concisely summarize which combinations of ACBC stages make sense.

One challenge in our analysis is that the *BYO* stage does not provide information about the outside good (i.e., the utility of the option not to buy). Consequently, we cannot predict choices in free-choice HOTs when solely using information provided in the *BYO* stage. Fortunately, in all our ACBC implementations, the configured ideal product of a participant was also part of the *Calibration* stage. This allows us to overcome this issue, by additionally utilizing the corresponding answers from the *Calibration* stage on the five-point purchase likelihood scale.

[15] We thank one reviewer for suggesting a more detailed analysis of the ACBC stages.

We "binarize" the answers such that we have one additional choice task in this modified *BYO* stage (values greater or equal to three = 1, 0 otherwise).

Dropping later choice tasks in CBC studies has recently been analyzed by Li et al. (2022). The authors investigate how the number of choice tasks (*"how much you ask"*) affects the predictive validity[16] and find that participants adapt their behavior to the specific (in-sample) tasks using task-specific decision processes ultimately leading to a reduced predictive validity. Given that we also have CBC conditions, and in the Food Processors study, even ones with and without incentive alignment, the analysis in this Web Appendix contributes by replicating and extending the analysis of Li et al. (2022), who only had a CBC study with incentive alignment. Furthermore, we can analyze whether ACBC suffers from the same problem (i.e., decision process adaptation) and whether later stages reduce predictive validity. We do not expect this to happen, because the adaptive design changes the type of choice tasks in each stage on purpose. Therefore, adaptive design in Sawtooth's ACBC should be less affected by boredom and adaptation strategies than regular CBC studies.[17] Nevertheless, it is unclear how incentive alignment affects the different stages of the ACBC.

Fig. H22 shows the resulting mean hit probabilities (MHP in %) for each study, condition, and HOT.[18] We restrict our attention to the first three HOTs for the Food Processor study. To combine CBC and ACBC in one plot, we align the results by the median time in seconds that the participants needed to finish the survey up to a particular choice task or stage (see, Toubia et al. 2012 for the idea on a time-equalizing analysis). For ACBC conditions, we either show

---

[16] Li et al. (2022) specifically focus on external validity by evaluating the predictive accuracy for hold-out tasks containing a larger number of alternatives, i.e., making predictions about choice behavior in different settings. We do the same in our hold-out tasks, but adopt a more neutral term, generally referring to it as "predictive validity."
[17] It is important to highlight that this argument holds for Sawtooth's ACBC. Other types of adaptive designs could suffer from the same problem as regular CBC studies.
[18] We plot 95% confidence intervals to indicate the uncertainty in our results. However, refrain from performing tests for all comparisons as this analysis has an explorative/descriptive character. Note that for some comparisons, future research should further study the robustness of the patterns found in our study.

three or four points (i.e., stages), depending on whether a HOT is free- or forced-choice (the HOT in the PlayStation 4 Bundles study and HOT 3 in the Food Processors study).

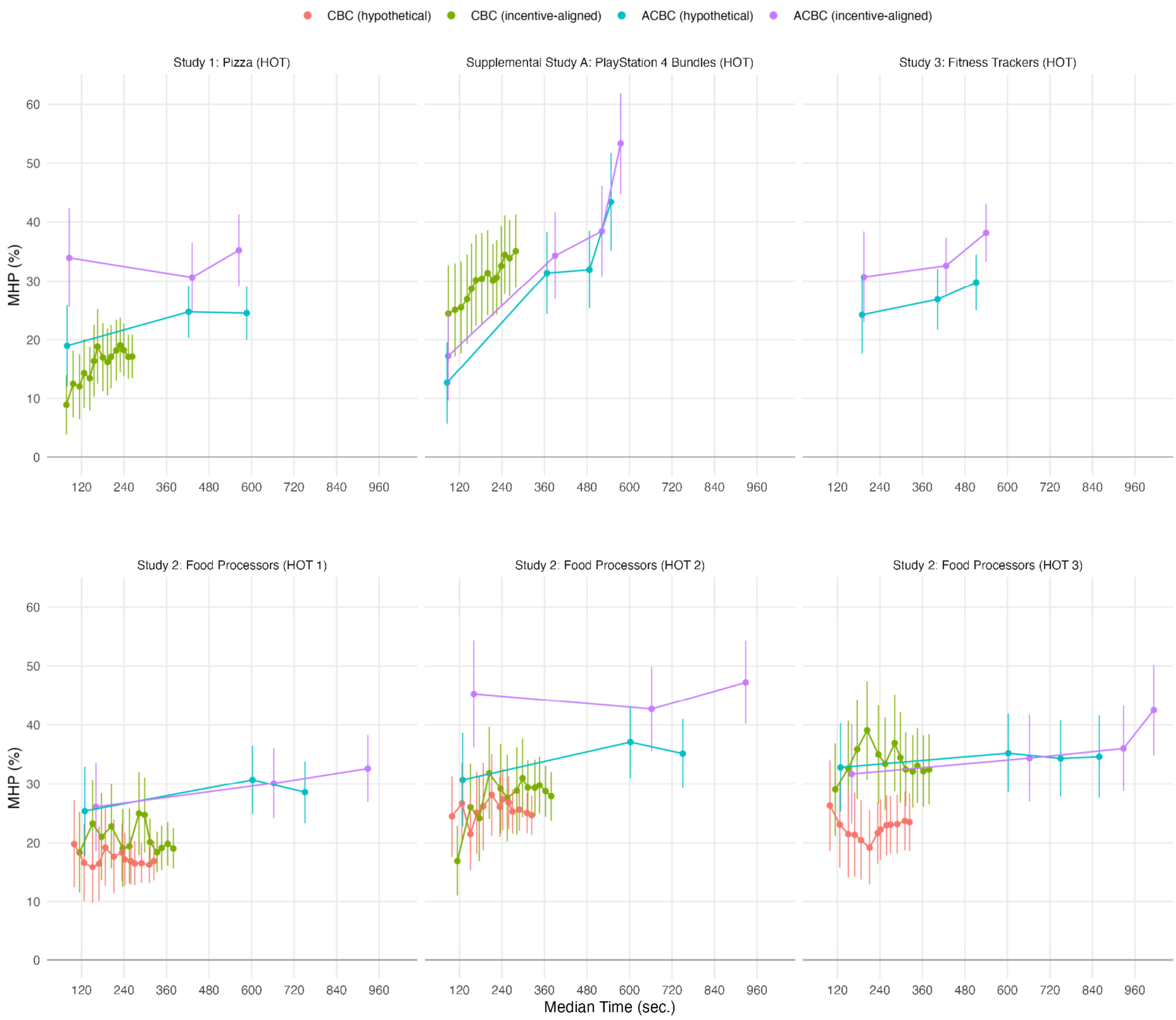


Fig. H22. Mean hit probabilities depending on the amount of information used in estimation

Several interesting patterns emerge: First, as already mentioned earlier in Web Appendix H, the total times of ACBCs are considerably longer than the corresponding CBC conditions (of the same study). Furthermore, even over time, we see that the general order of the results holds: ACBC conditions outperform CBC conditions, and also conditions with incentive alignment have higher MHP than hypothetical ones. In some cases (Study 1 or HOT 2 in Study 2), ACBC conditions even outperform CBC conditions after the *BYO* stage already. Exceptions to the general order of results are (a) Suppl. Study A and Study 2 (HOT 3), where the incentive-aligned CBCs outperform the ACBC counterparts until some point, (b) HOT 1 and HOT 3 in Study 2, where the hypothetical ACBC is just slightly better than the incentive-aligned ACBC until the *Tournament* stage, and (c) in the same study in HOT 1 and 2, hypothetical CBC outperforms the incentive-aligned CBC until the second or third choice tasks, respectively. Second, we find the same pattern as Li et al. (2022) for the CBC conditions, i.e., the maximum predictive validity is not achieved when using all choice tasks. In Study 2, the hypothetical CBC condition hardly improves after two choice tasks for all three HOTs. Third, the pattern of Li et al. (2022) does not emerge for the ACBC conditions with incentive alignment. Here, the highest predictive validity is always achieved in the third (or fourth) stage. Interestingly, the hypothetical ACBC in Study 2, does not improve after the *Screening* stage. We do find the following results for the question of which stage has the largest increases in MHP due to incentive alignment? Study 1, HOT 2 in Study 2, and Study 3 suggest that the *BYO* (+39.8%) and *Tournament* (+30.1) stages benefit most from incentive alignment. In free-choice HOTs (Supplemental Study A and HOT 3 in Study 2), the *BYO* (+16.5%) and the *Calibration* (+22.8%) stages have stronger effects. That the *Tournament* stage benefits from incentive alignment in Studies 1, Supplemental Study A, and Study 2 was expected, as specifically this stage was incentivized (*Choice Tournament* winner had a 50% chance to be realized for the incentivization). However, the noteworthy finding that other stages also gain advantages from incentive alignment suggests that participants comprehended the interlinked nature of the ACBC process. This finding is also consistent with the time spent per choice task in the Web Appendix *"Comparison of survey times of all stages in ACBC studies."* While we observe that incentive alignment significantly increases deliberation across all studies, the primary driver for this result is the increased deliberation in the *BYO* and *Screening* stages (refer to Fig. H15, H16, and H19).

We conclude that *"the more you ask, the less you get",* as Li et al. (2022) write in the title of their paper, also holds for our CBC conditions, but not in the case of adaptive designs, especially when ACBC is combined with incentive alignment. Here we do not find a systematic decline in predictive validity over time. Asking more is beneficial.

References

Abernethy, J., Evgeniou, T., Toubia, O., & Vert, J.-P. (2005). *An optimization framework for the adaptive design of robust choice questionnaires*. http://www.columbia.edu/~ot2107/rnq.pdf.

Abernethy, J., Evgeniou, T., Toubia, O., & Vert, J.-P. (2008). Eliciting consumer preferences using robust adaptive choice questionnaires. *IEEE Transactions on Knowledge and Data Engineering, 20*, 145–155. doi:10.1109/TKDE.2007.190632.

Akinc, D., & Vandebroek, M. (2018). Bayesian estimation of mixed logit models: Selecting an appropriate prior for the covariance matrix. *Journal of Choice Modelling, 29*, 133–151. doi:10.1016/j.jocm.2017.11.004.

Allenby, G. M., Brazell, J. D., Howell, J. R., & Rossi, P. E. (2014). Economic valuation of product features. *Quantitative Marketing and Economics, 12*, 421–456. doi:10.1007/s11129-014-9150-x.

Allenby, G. M., & Ginter, J. L. (1995). Using extremes to design products and segment markets. *Journal of Marketing Research, 32*, 392–403. doi:10.1177/002224379503200402.

Allenby, G. M., & Rossi, P. E. (1998). Marketing models of consumer heterogeneity. *Journal of Econometrics, 89*, 57–78. doi:10.1016/S0304-4076(98)00055-4.

Almario, C. V., Eberlein, S., Khalil, C., & Spiegel, B. M. R. (2021). Determining patient treatment preferences for management of acute pain episodes in irritable bowel syndrome. *Neurogastroenterology and motility: the official journal of the European Gastrointestinal Motility Society*, e14145. doi:10.1111/nmo.14145.

Al-Omari, B. (2017). Patient preferences for the pharmacological treatment of osteoarthritis using adaptive choice-based conjoint (ACBC) analysis: A pilot study. *European Journal for Person Centered Healthcare, 5*, 220–224. doi:10.5750/ejpch.v5i2.1293.

Al-Omari, B., & McMeekin, P. (2020). Patients' preferences regarding osteoarthritis medications: An adaptive choice-based conjoint analysis study. *Patient preference and adherence, 14*, 2501–2515. doi:10.2147/PPA.S283922.

Al-Omari, B., Sim, J., Croft, P., & Frisher, M. (2017). Generating individual patient preferences for the treatment of osteoarthritis using Adaptive Choice-Based Conjoint (ACBC) analysis. *Rheumatology and Therapy, 4*, 167–182. doi:10.1007/s40744-017-0056-4.

Bacon, L., & Lenk, P. (2012). Augmenting discrete-choice data to identify common preference scales for inter-subject analyses. *Quantitative Marketing and Economics, 10*, 453–474. doi:10.1007/s11129-012-9124-9.

Baier, D., & Brusch, M. (2021). *Conjointanalyse*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Banerjee, S., Al-Qaheri, H., Chiş, M., & Ali Hassan, M. B. (2011). An adaptive questionnaire generation using learning from fuzzy and post clustering of customers responses: An experience with communication products. *Telecommunication Systems, 46*, 273–284. doi:10.1007/s11235-010-9290-6.

Barnard, J., McCulloch, R., & Meng, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica, 10*(4), 1281–1311.

Bauer, R., Menrad, K., & Decker, T. (2015). Adaptive hybrid methods for choice-based conjoint analysis: A comparative study. *International Journal of Marketing Studies, 7*, 1–14. doi:10.5539/ijms.v7n1p1.

Becker, G. M., Degroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science, 9*, 226–232. doi:10.1002/bs.3830090304.

Bertsimas, D., & O'Hair, A. (2013). Learning preferences under noise and loss aversion: An optimization approach. *Operations Research, 61*, 1190–1199. doi:10.1287/opre.2013.1209.

Bijmolt, T. H., Van Heerde, H. J., & Pieters, R. G. (2005). New empirical generalizations on the determinants of price elasticity. *Journal of Marketing Research, 42*, 141–156. doi:10.1509/jmkr.42.2.141.62296.

Boesch, I. (2012). Preferences for attributes of industrial potatoes: An empirical investigation of the Swiss market. *American Journal of Potato Research, 89*, 199–206. doi:10.1007/s12230-012-9244-5.

Boesch, I., Schwaninger, M., Weber, M., & Scholz, R. W. (2013). Enhancing validity and reliability through feedback-driven exploration: A study in the context of conjoint analysis. *Systemic Practice and Action Research, 26*, 217–238. doi:10.1007/s11213-012-9248-6.

Böhm, S., Constantine, P., & Igler, B. (2013). Assessing user preferences for mobile applications in public transportation: A pre-study using a conjoint-based research methodology. *International Journal of Business and Management Studies, 5*(1), 485–494.

Boyer, R. H., Hunka, A. D., Linder, M., Whalen, K. A., & Habibi, S. (2021a). Product labels for the circular economy: Are customers willing to pay for circular? *Sustainable Production and Consumption, 27*, 61–71. doi:10.1016/j.spc.2020.10.010.

Boyer, R. H. W., Hunka, A. D., & Whalen, K. A. (2021b). Consumer demand for circular products: Identifying customer segments in the circular economy. *Sustainability, 13*, 12348. doi:10.3390/su132212348.

Brand, B. M., & Baier, D. (2020). Adaptive CBC: Are the benefits justifying its additional efforts compared to CBC? *Archives of Data Science, Series A, 6*, 1–22. doi:10.5445/KSP/1000098011/06.

Brand, B. M., & Rausch, T. M. (2021). Examining sustainability surcharges for outdoor apparel using Adaptive Choice-Based Conjoint analysis. *Journal of Cleaner Production, 289*, 125654. doi:10.1016/j.jclepro.2020.125654.

Brand, B. M., Rausch, T. M., & Brandel, J. (2022). The importance of sustainability aspects when purchasing online: comparing Generation X and Generation Z. *Sustainability, 14*, 5689. doi:10.3390/su14095689.

Brazell, J. D., Diener, C. G., Karniouchina, E., Moore, W. L., Séverin, V., & Uldry, P.-F. (2006). The no-choice option and dual response choice designs. *Marketing Letters, 17*, 255–268. doi:10.1007/s11002-006-7943-8.

Brodock, J. L., Hayes, J. E., Masterson, T. D., & Hopfer, H. (2021). Differences in preferred fat level, sweetener type, and amount of added sugar in chocolate milk in a choice task relate to physical activity and orthorexia. *Appetite, 163*, 105214. doi:10.1016/j.appet.2021.105214.

Bruno, H. A., Cebollada, J., & Chintagunta, P. K. (2018). Targeting Mr. or Mrs. Smith: Modeling and leveraging intrahousehold heterogeneity in brand choice behavior. *Marketing Science, 37*, 631–648. doi:10.1287/mksc.2018.1088.

Buoye, A., Williams, L., & Weiner, J. (2018). WAR: What else is it good for? A comparison of maximum difference scaling, adaptive choice-based conjoint, constant sum scaling and wallet allocation rule utilities. *Journal of Creating Value, 4*, 61–81. doi:10.1177/2394964318771794.

Burke, P. F., Burton, C., Huybers, T., Islam, T., Louviere, J. J., & Wise, C. (2010). The scale-adusted latent class model: application to museum visitation. *Tourism Analysis, 15*, 147–165. doi:10.3727/108354210X12724863327605.

Burs, L., Roemer, E., Worm, S., & Masini, A. (2020). Are they all equal? Uncovering adopter groups of battery electric vehicles. *Sustainability, 12*, 2815. doi:10.3390/su12072815.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*(1), 1-32.

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., et al. (2022). *shiny: Web Application Framework for R* .

Chapman, C. N., Alford, J. L., Johnson, C., & Weidemann, R. (2009). CBC vs. ACBC: Comparing results with real product selection. https://www.sawtoothsoftware.com/support/technical-papers/adaptive-cbc-papers/cbc-vs-acbc-comparing-results-with-real-product-selection-2009.

Chassot, S., Hampl, N., & Wüstenhagen, R. (2014). When energy policy meets free-market capitalists: The moderating influence of worldviews on risk perception and renewable energy investment decisions. *Energy Research & Social Science, 3*, 143–151. doi:10.1016/j.erss.2014.07.013.

Chowdhury, M., Salam, K., & Tay, R. (2016). Consumer preferences and policy implications for the green car market. *Marketing Intelligence & Planning, 34*, 810–827. doi:10.1108/MIP-08-2015-0167.

Cocquyt, A., Crucke, S., & Slabbinck, H. (2020). Organizational characteristics explaining participation in sustainable business models in the sharing economy: Evidence from the fashion industry using conjoint analysis. *Business Strategy and the Environment, 29*, 2603–2613. doi:10.1002/bse.2523.

Consonni, G., Deldossi, L., & Saggini, E. (2020). Accounting for model uncertainty in individualized designs for discrete choice experiments. *Journal of Quality Technology, 52*, 81–96. doi:10.1080/00224065.2019.1569962.

Cook, N., Geier, A., Schmid, A., Hirschfield, G., Kautz, A., Schattenberg, J. M., et al. (2019). The patient perspectives on future therapeutic options in NASH and patient needs. *Frontiers in Medicine, 6*, 1–14. doi:10.3389/fmed.2019.00061.

Cook, N., Geier, A., Schmid, A., Hirschfield, G., Kautz, A., Schattenberg, J. M., et al. (2020). Assessing physician preferences on future therapeutic options and diagnostic practices in non-alcoholic steatohepatitis. *JHEP reports : innovation in hepatology, 2*, 100081. doi:10.1016/j.jhepr.2020.100081.

Côté, E., & Salm, S. (2022). Risk-adjusted preferences of utility companies and institutional investors for battery storage and green hydrogen investment. *Energy Policy, 163*, 112821. doi:10.1016/j.enpol.2022.112821.

Crabbe, M., Akinc, D., & Vandebroek, M. (2014). Fast algorithms to generate individualized designs for the mixed logit choice model. *Transportation Research Part B: Methodological, 60*, 1–15. doi:10.1016/j.trb.2013.11.008.

Croissant, Y. (2020). Estimation of random utility nodels in R: The mlogit package. *Journal of Statistical Software, 95*, 1–41. doi:10.18637/jss.v095.i11.

Cunningham, C. E., Chen, Y., Vaillancourt, T., Rimas, H., Deal, K., Cunningham, L. J., et al. (2015). Modeling the anti-cyberbullying preferences of university students: Adaptive choice-based conjoint analysis. *Aggressive Behavior, 41*, 369–385. doi:10.1002/ab.21560.

Cunningham, C. E., Deal, K., & Chen, Y. (2010). Adaptive choice-based conjoint analysis: A new patient-centered approach to the assessment of health service preferences. *The Patient: Patient-Centered Outcomes Research, 3*, 257–273. doi:10.2165/11537870-000000000-00000.

Currim, I. S., Weinberg, C. B., & Wittink, D. R. (1981). Design of subscription programs for a performing arts series. *Journal of Consumer Research, 8*, 67–75. doi:10.1086/208842.

Daly, A., Hess, S., & Train, K. (2012). Assuring finite moments for willingness to pay in random coefficient models. *Transportation, 39*, 19–31. doi:10.1007/s11116-011-9331-3.

Danthurebandara, V., Yu, J., & Vandebroek, M. (2015). Designing choice experiments by optimizing the complexity level to individual abilities. *Quantitative Marketing and Economics, 13*, 1–26. doi:10.1007/s11129-014-9152-8.

Danthurebandara, V. M., Yu, J., & Vandebroek, M. (2011). Sequential choice designs to estimate the heterogeneity distribution of willingness-to-pay. *Quantitative Marketing and Economics, 9*, 429–448. doi:10.1007/s11129-011-9106-3.

de Groot, I. B., Otten, W., Dijs-Elsinga, J., Smeets, H. J., Kievit, J., & Marang-van de Mheen, P J. (2012). Choosing between hospitals: the influence of the experiences of other patients. *Medical Decision Making, 32*, 764–778. doi:10.1177/0272989X12443416.

de Groot, I. B., Otten, W., Smeets, H. J., & Marang-van de Mheen, P. J. (2011). Is the impact of hospital performance data greater in patients who have compared hospitals? *BMC Health Services Research, 11*, 214–224. doi:10.1186/1472-6963-11-214.

Deal, K. (2014). Segmenting patients and physicians using preferences from discrete choice experiments. *The Patient: Patient-Centered Outcomes Research, 7*, 5–21. doi:10.1007/s40271-013-0037-9.

Del Toro-Gipson, R. S., Rizzo, P. V., Hanson, D. J., & Drake, M. A. (2021). Consumer perception of smoked Cheddar cheese. *Journal of Dairy Science, 104*, 1560–1575. doi:10.3168/jds.2020-18711.

Desarbo, W. S., Ramaswamy, V., & Cohen, S. H. (1995). Market segmentation with choice-based conjoint analysis. *Marketing Letters, 6*, 137–147. doi:10.1007/BF00994929.

Deshwal, A., Kumar, S., Tripathi, P., & Singh, V. (2022). To determine the market share of marketing channels involving in marketing of tomato seed (VNR) in Meerut district of Uttar Pradesh. *The Pharma Innovation Journal, 11*(5S), 729–731.

Ding, M. (2007). An incentive-aligned mechanism for conjoint analysis. *Journal of Marketing Research, 44*, 214–223. doi:10.1509/jmkr.44.2.214.

Ding, M., Grewal, R., & Liechty, J. C. (2005). Incentive-aligned conjoint analysis. *Journal of Marketing Research, 42*, 67–82. doi:10.1509/jmkr.42.1.67.56890.

Ding, M., Park, Y.-H., & Bradlow, E. T. (2009). Barter markets for conjoint analysis. *Management Science, 55*, 1003–1017. doi:10.1287/mnsc.1090.1003.

Dong, S., Ding, M., & Huber, J. (2010). A simple mechanism to incentive-align conjoint experiments. *International Journal of Research in Marketing, 27*, 25–32. doi:10.1016/j.ijresmar.2009.09.004.

Dzyabura, D., & Hauser, J. R. (2011). Active machine learning for consideration heuristics. *Marketing Science, 30*, 801–819. doi:10.1287/mksc.1110.0660.

Ebbers, F., Zibuschka, J., Zimmermann, C., & Hinz, O. (2021). User preferences for privacy features in digital assistants. *Electronic Markets, 31*, 411–426. doi:10.1007/s12525-020-00447-y.

Eggers, F. (2008). Adaptive design techniques for complex choice-based conjoint experiments. *Marketing ZFP, 30*, 67–75. doi:10.15358/0344-1369-2008-JRM-2-67.

Eggers, F. (2012). Would you like some more?: Voluntary choice sets and progressive incentive alignment in conjoint analysis. In S. Bharadwaj & J. Hulland (Eds.), *Marketing Theory and Applications*: *2012 AMA Winter Educators' Conference* (pp. 381–382). United States: American Marketing Association.

Eggers, F., & Sattler, H. (2009). Hybrid individualized two-level choice-based conjoint (HIT-CBC): A new method for measuring preference structures with many attribute levels. *International Journal of Research in Marketing, 26*, 108–118. doi:10.1016/j.ijresmar.2009.01.002.

Ellickson, P. B., Lovett, M. J., & Ranjan, B. (2019). Product launches with new attributes: a hybrid conjoint–consumer panel technique for estimating demand. *Journal of Marketing Research, 56*, 709–731. doi:10.1177/0022243719843132.

Faber, I. (2015). Evaluating the military retirement system: An adaptive choice-based conjoint analysis approach. *SSRN Electronic Journal*.

Feilhauer, E., Schnitzer, M., Walde, J., & Tappeiner, G. (2022). What residents of potential Olympic cities want: using conjoint analysis to deal with dominant and heterogeneous preferences. *Current Issues in Tourism*, 1–14. doi:10.1080/13683500.2022.2067030.

Flöthmann, C., Hoberg, K., & Wieland, A. (2018). Competency requirements of supply chain planners & analysts and personal preferences of hiring managers. *Supply Chain Management: An International Journal, 23*, 480–499. doi:10.1108/SCM-03-2018-0101.

Frischknecht, B. D., Eckert, C., Geweke, J., & Louviere, J. J. (2014). A simple method for estimating preference parameters for individuals. *International Journal of Research in Marketing, 31*, 35–48. doi:10.1016/j.ijresmar.2013.07.005.

Fuchs, M., & Hovemann, G. (2022). Consumer preferences for circular outdoor sporting goods: An Adaptive Choice-Based Conjoint analysis among residents of European outdoor markets. *Cleaner Engineering and Technology, 11*, 100556. doi:10.1016/j.clet.2022.100556.

Gabry, J., & Mahr, T. (2022). bayesplot: plotting for Bayesian models. R package version 1.10.0. https://mc-stan.org/bayesplot/.

Gaskin, S. P., Evgeniou, T., Bailiff, D., & Hauser, J. R. (2007). Two-stage models: Identifying non-compensatory heuristics for the consideration set then adaptive polyhedral methods within the consideration set. https://www.semanticscholar.org/paper/Two-Stage-Models-%3A-Identifying-Non-Compensatory-for-Gaskin-Evgeniou/c7df6043edcdfe9bad9ab0c676acf0419fb3865b.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayestian data analysis* (3rd ed.). United States: Chapman & Hall/CRC Boca Raton.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–511.

Gensler, S., Hinz, O., Skiera, B., & Theysohn, S. (2012). Willingness-to-pay estimation with choice-based conjoint analysis: Addressing extreme response behavior with individually adapted designs. *European Journal of Operational Research, 219*, 368–378. doi:10.1016/j.ejor.2012.01.002.

Giessmann, A., Feurer, S., Schneider, T., & Wenzel, S. (2015). Consumers' preferences on platform as a service and SAP's engineering response. *Econstor*, https://www.econstor.eu/handle/10419/106129.

Gundlach, H., & Hofmann, U. (2020). Information search, behavioural economics, and relevance decisions in the online media industry: how strongly do the algorithms of intermediaries influence the relevance evaluation of information? *Journal of Media Business Studies*, 1–20. doi:10.1080/16522354.2020.1854602.

Gutierrez, J. I., Dubov, A., Altice, F. L., & Vlahov, D. (2021). Preferences for pre-exposure prophylaxis among U.S. military men who have sex with men: results of an adaptive choice based conjoint analysis study. *Military Medical Research, 8*, 32. doi:10.1186/s40779-021-00323-6.

Gutierrez, J. I., Vlahov, D., Dubov, A., & Altice, F. L. (2022). Preferences for long-acting and alternative modalities for prep among military men who have sex with men: Segmentation results of an adaptive choice-based conjoint analysis study. *Journal of urban health : bulletin of the New York Academy of Medicine, 99*, 277–292. doi:10.1007/s11524-022-00615-9.

Habann, F., Zerres, C., & Zaworski, L. (2018). Investigating the segment specific preferences for hedonic and utilitarian online-shop characteristics: The case of German online wine shops. *International Journal of Internet Marketing and Advertising, 12*(3), 255–269.

Haggenmüller, S., Krieghoff-Henning, E., Jutzi, T., Trapp, N., Kiehl, L., Utikal, J. S., et al. (2021). Digital natives' preferences on mobile artificial intelligence apps for skin cancer diagnostics: Survey study. *JMIR mHealth and uHealth, 9*, e22909. doi:10.2196/22909.

Halme, M., & Kallio, M. (2014). Likelihood estimation of consumer preferences in choice-based conjoint analysis. *European Journal of Operational Research, 239*, 556–564. doi:10.1016/j.ejor.2014.05.044.

Harwood, W. S., & Drake, M. (2018). Identification and characterization of fluid milk consumer groups. *Journal of Dairy Science, 101*, 8860–8874. doi:10.3168/jds.2018-14855.

Harwood, W. S., & Drake, M. (2019). Understanding implicit and explicit consumer desires for protein bars, powders, and beverages. *Journal of Sensory Studies, 34*, e12493. doi:10.1111/joss.12493.

Harwood, W. S., & Drake, M. (2020). Validation of fluid milk consumer segments using qualitative multivariate analysis. *Journal of Dairy Science, 103*, 10036–10047. doi:10.3168/jds.2019-17797.

Hauser, J. R., Eggers, F., & Selove, M. (2019). The strategic implications of scale in choice-based conjoint analysis. *Marketing Science, 38*, 913–1084. doi:10.1287/mksc.2019.1178.

Hedges, L. V., & Olkin, I. (1985). *Statistical method for meta-analysis*. Orlando, FL, US: Academic Press Inc.

Heidel, A., Hagist, C., & Schlereth, C. (2021). Pricing through health apps generated data-Digital dividend as a game changer: Discrete choice experiment. *PLoS ONE, 16*, e0254786. doi:10.1371/journal.pone.0254786.

Henderson, T., & Liu, Q. (2017). Efficient design and analysis for a selective choice process. *Journal of Marketing Research, 54*, 430–446. doi:10.1509/jmr.13.0545.

Hinnen, G., Hille, S. L., & Wittmer, A. (2017). Willingness to pay for green products in air travel: Ready for take-off? *Business Strategy and the Environment, 26*, 197–208. doi:10.1002/bse.1909.

Hoffman, M. D., & Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research, 15*(1), 1593–1623.

Holopainen, J., Mattila, O., Pöyry, E., & Parvinen, P. (2020). Applying design science research methodology in the development of virtual reality forest management services. *Forest Policy and Economics, 116*, 102190. doi:10.1016/j.forpol.2020.102190.

Hoppe, D., Keller, H., & Horstmann, F. (2022). Got employer image? How applicants choose their employer. *Corporate Reputation Review, 25*, 139–159. doi:10.1057/s41299-021-00119-3.

Howell, J. R., Ebbes, P., & Liechty, J. C. (2021). Gremlins in the data: Identifying the information content of research subjects. *Journal of Marketing Research, 58*, 74–94. doi:10.1177/0022243720965930.

Hu, X., Lima, M. F., McLean, R., & Sun, Z. (2022). Exploring preferences for biodiversity and wild parks in Chinese cities: A conjoint analysis study in Hangzhou. *Urban Forestry & Urban Greening, 73*, 127595. doi:10.1016/j.ufug.2022.127595.

Huang, D., & Luo, L. (2016). Consumer preference elicitation of complex products using fuzzy support vector machine active learning. *Marketing Science, 35*, 445–464. doi:10.1287/mksc.2015.0946.

Huber, J., & Train, K. (2001). On the similarity of classical and Bayesian estimates of individual mean partworths. *Marketing Letters, 12*(3), 259–269.

Hunka, A. D., Linder, M., & Habibi, S. (2021). Determinants of consumer demand for circular economy products. A case for reuse and remanufacturing for sustainable development. *Business Strategy and the Environment, 30*, 535–550. doi:10.1002/bse.2636.

Inci, E., Tatar Taspinar, Z., & Ulengin, B. (2022). A choice experiment on preferences for electric and hybrid cars in Istanbul. *Transportation Research Part D: Transport and Environment, 107*, 103295. doi:10.1016/j.trd.2022.103295.

Isihara, P., Shi, C., Ward, J., O'Malley, L., Laney, S., Diedrichs, D., et al. (2020). Identifying most typical and most ideal attribute levels in small populations of expert decision makers: Studying the Go/No Go decision of disaster relief organizations. *Journal of Choice Modelling, 35*, 100204. doi:10.1016/j.jocm.2020.100204.

Issa, A. M., Tufail, W., Atehortua, N., & McKeever, J. (2013). A national study of breast and colorectal cancer patients' decision-making for novel personalized medicine genomic diagnostics. *Personalized Medicine, 10*, 245–256. doi:10.2217/pme.13.17.

Jazdzik-Osmolska, A. (2021). Willingness to pay for road safety improvements in Poland. *European Research Studies Journal, 24*, 96–117. doi:10.35808/ersj/2462.

Jedidi, K., & Zhang, Z. J. (2002). Augmenting conjoint analysis to estimate consumer reservation price. *Management Science, 48*, 1350–1368. doi:10.1287/mnsc.48.10.1350.272.

Jervis, M. G., Jervis, S. M., Guthrie, B., & Drake, M. (2014a). Determining children's perceptions, opinions and attitudes for sliced sandwich breads. *Journal of Sensory Studies, 29*, 351–361. doi:10.1111/joss.12116.

Jervis, S. M., Ennis, J. M., & Drake, M. (2012). A comparison of adaptive choice-based conjoint and choice-based conjoint to determine key choice attributes of sour cream with limited sample size. *Journal of Sensory Studies, 27*, 451–462. doi:10.1111/joss.12009.

Jervis, S. M., Jervis, M. G., Guthrie, B., & Drake, M. (2014b). The Efficacy of using photographs to represent attributes of sliced sandwich bread in an adaptive choice-based conjoint. *Journal of Sensory Studies, 29*, 64–73. doi:10.1111/joss.12082.

Johnson, R. M. (2000). Monotonicity constraints in choice-based conjoint with Hierarchical Bayes. https://sawtoothsoftware.com/resources/technical-papers/monotonicity-contraints-in-choice-based-conjoint-with-hierarchical-bayes.

Johnson, R. M., & Orme, B. K. (2007). A new approach to adaptive CBC. Sawtooth Software Inc. http://www.sawtoothsoftware.com/support/technical-papers/adaptive-cbc-papers/a-new-approach-to-adaptive-cbc-2007. Accessed 18 June 2014.

Joo, M., Thompson, M. L., & Allenby, G. M. (2019). Optimal product design by sequential experiments in high dimensions. *Management Science, 65*, 3235–3254. doi:10.1287/mnsc.2018.3088.

Joo, W., Almario, C. V., Ishimori, M., Park, Y., Jusufagic, A., Noah, B., et al. (2020). Examining treatment decision-making among patients with axial spondyloarthritis: Insights from a conjoint analysis survey. *ACR open rheumatology, 2*, 391–400. doi:10.1002/acr2.11151.

Kamakura, W. A., & Russell, G. J. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research, 26*, 379–390. doi:10.1177/002224378902600401.

Kang, N., Ren, Y., Feinberg, F., & Papalambros, P. Y. (2019). *Form + function: Optimizing aesthetic product design via adaptive, geometrized preference elicitation.* (arXiv preprint arXiv:1912.05047.). https://arxiv.org/pdf/1912.05047.pdf.

Khan, R. A. J., Thaheem, M. J., & Ali, T. H. (2020). Are Pakistani homebuyers ready to adopt sustainable housing? An insight into their willingness to pay. *Energy Policy, 143*, 111598. doi:10.1016/j.enpol.2020.111598.

Kim, M. K., Lopetcharat, K., & Drake, M. (2013). Influence of packaging information on consumer liking of chocolate milk. *Journal of Dairy Science, 96*, 4843–4856. doi:10.3168/jds.2012-6399.

Klein, G., Shtudiner, Z., & Zwilling, M. (2021). Why do peer-to-peer (P2P) lending platforms fail? The gap between P2P lenders' preferences and the platforms' intentions. *Electronic Commerce Research*. doi:10.1007/s10660-021-09489-6.

Klein, M., Linden, E., & Wittmer, A. (2019). Influence of marketing instruments on consumer behavior in the process of purchasing leisure flight tickets. *Marketing Review St. Gallen, 3*(May), 30–38.

Knoerl, R., Berry, D., Meyerhardt, J. A., Reyes, K., Salehi, E., Thornton, K., et al. (2022). Identifying participants' preferences for modifiable chemotherapy-induced peripheral neuropathy prevention clinical trial factors: an adaptive choice-based conjoint analysis. *Supportive Care in Cancer, 30*, 9963–9973. doi:10.1007/s00520-022-07447-y.

Kouki-Block, M., & Wellbrock, C.-M. (2021). Influenced by media brands? A conjoint experiment on the effect of media brands on online media planners' decision-making. *Journal of Media Business Studies*, 1–23. doi:10.1080/16522354.2021.1899741.

Krewson, C. N., & Owens, R. J. (2022). Judicial philosophy and the public's support for courts. *Political Research Quarterly*, 106591292211197. doi:10.1177/10659129221119741.

Lagerkvist, C. J., Okello, J., & Karanja, N. (2012). Anchored vs. relative best–worst scaling and latent class vs. hierarchical Bayesian analysis of best–worst choice data: Investigating the importance of food quality attributes in a developing country. *Food Quality and Preference, 25*, 29–40. doi:10.1016/j.foodqual.2012.01.002.

Lattery, K. (2010). Anchoring maximum difference scaling against a threshold–dual response and direct binary responses. In Sawtooth Software (Ed.), *Sawtooth Software Conference, Newport Beach, California, 06.10.-08.10.* (pp. 91–106).

Lehmann, N., Sloot, D., Ardone, A., & Fichtner, W. (2022). Willingness to pay for regional electricity generation – A question of green values and regional product beliefs? *Energy Economics, 110*, 106003. doi:10.1016/j.eneco.2022.106003.

Lemken, D. (2021). The price penalty for red meat substitutes in popular dishes and the diversity in substitution. *PLoS ONE, 16*, e0252675. doi:10.1371/journal.pone.0252675.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis, 100*, 1989–2001. doi:10.1016/j.jmva.2009.04.008.

Li, Y., Krefeld-Schwalb, A., Wall, D. G., Johnson, E. J., Toubia, O., & Bartels, D. M. (2022). The more you ask, the less you get: When additional questions hurt external validity. *Journal of Marketing Research, 59*, 963-982. doi:10.1177/00222437211073581.

Lima, M. F., Ward Thompson, C., & Aspinall, P. (2020). Friendly communities and outdoor spaces in contexts of urban population decline. *Land, 9*, 439. doi:10.3390/land9110439.

Lima, M. F., Ward Thompson, C., Aspinall, P., & Bell, S. (2022). Communities facing urban depopulation: exploring people's environmental preferences. A case study of Lisbon, Portugal. *Cities & Health, 6*, 288–308. doi:10.1080/23748834.2020.1727820.

Louviere, J. J., Lings, I., Islam, T., Gudergan, S., & Flynn, T. (2013). An introduction to the application of (case 1) best-worst scaling in marketing research. *International Journal of Research in Marketing, 30*, 292–303. doi:10.1016/j.ijresmar.2012.10.002.

Louviere, J. J., & Woodworth, G. G. (1991a). *Best-worst scaling: A model for the largest difference judgments*. Working Paper .

Louviere, J. J., & Woodworth, G. G. (1991b). *Best-worst scaling: A model for the largest difference judgments*. Working Paper .

Lysaght, T., Ballantyne, A., Toh, H. J., Lau, A., Ong, S., Schaefer, O., et al. (2021). Trust and trade-offs in sharing data for precision medicine: A national survey of Singapore. *Journal of personalized medicine*. doi:10.3390/jpm11090921.

Mai, L., Köchling, A., & Wehner, M. C. (2022). "This student needs to stay back": To what degree would instructors rely on the recommendation of learning analytics? *SN computer science, 3*, 259. doi:10.1007/s42979-022-01137-6.

McCarthy, K. S., Parker, M., Ameerally, A., Drake, S. L., & Drake, M. A. (2017). Drivers of choice for fluid milk versus plant-based alternatives: What are consumer perceptions of fluid milk? *Journal of Dairy Science, 100*, 6125–6138. doi:10.3168/jds.2016-12519.

McLean, K. G., Hanson, D. J., Jervis, S. M., & Drake, M. A. (2017). Consumer perception of retail pork bacon attributes using adaptive choice-based conjoint analysis and maximum differential scaling. *Journal of Food Science, 82*, 2659–2668. doi:10.1111/1750-3841.13934.

McShane, B. B., & Böckenholt, U. (2017). Single paper meta-analysis: Benefits for study summary, theory-testing, and replicability. *Journal of Consumer Research, 43*, 1048-1063. doi:10.1093/jcr/ucw085.

Melliger, M., & Lilliestam, J. (2021). Effects of coordinating support policy changes on renewable power investor choices in Europe. *Energy Policy, 148*, 111993. doi:10.1016/j.enpol.2020.111993.

Mengelkamp, E., Schönland, T., Huber, J., & Weinhardt, C. (2019). The value of local electricity - A choice experiment among German residential customers. *Energy Policy, 130*, 294–303. doi:10.1016/j.enpol.2019.04.008.

Miaoulis, G., Parsons, H., & Free, V. (1990). Turf: A new planning approach for product line extensions. *Marketing Research, 2*(1), 28–40.

Mihale-Wilson, A. C., Zibuschka, J., & Hinz, O. (2019). User preferences and willingness to pay for in-vehicle assistance. *Electronic Markets, 29*, 37–53. doi:10.1007/s12525-019-00330-5.

Miller, K. M., Hofstetter, R., Krohmer, H., & Zhang, Z. J. (2011). How should consumers' willingness to pay be measured? An empirical comparison of state-of-the-art approaches. *Journal of Marketing Research, 48*, 172–184. doi:10.1509/jmkr.48.1.172.

Naous, D., Bonner, M., Humbert, M., & Legner, C. (2022). Learning from the past to improve the future. *Business & Information Systems Engineering, 64*, 597–614. doi:10.1007/s12599-022-00742-2.

Natter, M., Mild, A., Wagner, U., & Taudes, A. (2008). Practice prize report —Planning new tariffs at tele.ring: The application and impact of an integrated segmentation, targeting, and positioning tool. *Marketing Science, 27*, 600–609. doi:10.1287/mksc.1070.0307.

Nickkar, A., & Lee, Y.-J. (2022). Willingness to pay for advanced safety features in vehicles: An adaptive choice-based conjoint analysis approach. *Transportation Research Record*, 036119812210770. doi:10.1177/03611981221077077.

Oltman, A. E., Jervis, S. M., & Drake, M. A. (2014). Consumer attitudes and preferences for fresh market tomatoes. *Journal of Food Science, 79*, S2091-7. doi:10.1111/1750-3841.12638.

Oltman, A. E., Lopetcharat, K., Bastian, E., & Drake, M. A. (2015). Identifying key attributes for protein beverages. *Journal of Food Science, 80*, S1383-90. doi:10.1111/1750-3841.12877.

Orme, B. K. (2009a). Anchored scaling in MaxDiff using dual response. https://sawtoothsoftware.com/resources/technical-papers/anchored-scaling-in-maxdiff-using-deal-response.

Orme, B. K. (2009b). Fine-tuning CBC and adaptive CBC questionnaires. https://www.sawtoothsoftware.com/support/technical-papers/adaptive-cbc-papers/fine-tuning-cbc-and-adaptive-cbc-questionnaires-2009.

Orme, B. K. (2019). *Consistency cutoffs to identify" bad" respondents in CBC, ACBC, and MaxDiff* (Sawtooth Software Research Paper Series).

https://sawtoothsoftware.com/resources/technical-papers/consistency-cutoffs-to-identify-bad-respondents-in-cbc-acbc-and-maxdiff.
Orme, B. K., & Johnson, R. M. (2008). Testing adaptive CBC: Shorter questionnaires and BYO vs. “most likelies”. https://www.sawtoothsoftware.com/support/technical-papers/adaptive-cbc-papers/testing-adaptive-cbc-shorter-questionnaires-and-byo-vs-most-likelies-2008.
Otter, T. (2007). HB-analysis for multi-format adaptive CBC. In Sawtooth Software (Ed.), *Proceedings of the Sawtooth Software Conference*. Santa Rosa (pp. 111–126). United States: Sawtooth Software Inc.
Pachali, M. J., Kurz, P., & Otter, T. (2020). How to generalize from a hierarchical model? *Quantitative Marketing and Economics*. doi:10.1007/s11129-020-09226-7.
Pallone, S. N., Frazier, P. I., & Henderson, S. G. (2017). Bayes-optimal entropy pursuit for active choice-based preference learning. *Cornell University*(February), https://arxiv.org/abs/1702.07694.
Petrovich, B., Hille, S. L., & Wüstenhagen, R. (2019). Beauty and the budget: A segmentation of residential solar adopters. *Ecological Economics, 164*, 1-11. doi:10.1016/j.ecolecon.2019.106353.
Pierce, J. P., Strong, D. R., Stone, M. D., Villaseñor, A., Dimofte, C. V., Leas, E. C., et al. (2020). Real-world exposure to graphic warning labels on cigarette packages in US smokers: The CASA randomized trial protocol. *Contemporary clinical trials, 98*, 106152. doi:10.1016/j.cct.2020.106152.
Pleger, L. E., Mertes, A., Rey, A., & Brüesch, C. (2020). Allowing users to pick and choose: A conjoint analysis of end-user preferences of public e-services. *Government Information Quarterly, 37*, 101473. doi:10.1016/j.giq.2020.101473.
Racette, C. M., & Drake, M. (2022). Consumer perception of natural hot-pepper cheeses. *Journal of Dairy Science, 105*, 2166–2179. doi:10.3168/jds.2021-20808.
Rao, V. R. (2014). *Applied conjoint analysis*. Berlin - Heidelberg: Springer.
Reinisch, M., Marschner, N., Otto, T., Korfel, A., Stoffregen, C., & Wöckel, A. (2021). Patient preferences: Results of a German adaptive choice-based conjoint analysis (market research study sponsored by Eli Lilly and Company) in patients on palliative treatment for advanced breast cancer. *Breast Care*, 1–9. doi:10.1159/000513139.
Ren, M. Y., & Scott, C. (2017). Adaptive questionnaires for direct identification of optimal product design. *Cornell University*(January), https://arxiv.org/abs/1701.01231.
Rheindorf, J., Hagist, C., & Schlereth, C. (2021). Call (and pay) the midwife: A discrete choice experiment on mothers’ preferences and their willingness to pay for midwifery care. *International Journal of Nursing Studies, 124*, 104096. doi:10.1016/j.ijnurstu.2021.104096.
Rivera, E., Timperio, A., Loh, V. H., Deforche, B., & Veitch, J. (2021). Important park features for encouraging park visitation, physical activity and social interaction among adolescents: A conjoint analysis. *Health & place, 70*, 102617. doi:10.1016/j.healthplace.2021.102617.
Rizzo, P. V., Harwood, W. S., & Drake, M. A. (2020). Consumer desires and perceptions of lactose-free milk. *Journal of Dairy Science, 103*, 6950–6966. doi:10.3168/jds.2019-17940.
Rohrmann, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie, 9*(1), 222–245.
Ronda, L., Abril, C., & Valor, C. (2021). Job choice decisions: understanding the role of nonnegotiable attributes and trade-offs in effective segmentation. *Management Decision, 59*, 1546–1561. doi:10.1108/MD-10-2019-1472.

Salm, S. (2017). The investor-specific price of renewable energy project risk: A choice experiment with incumbent utilities and institutional investors. *Renewable and Sustainable Energy Reviews, 82*, 1364–1375. doi:10.1016/j.rser.2017.04.009.
Salm, S., Hille, S. L., & Wüstenhagen, R. (2016). What are retail investors' risk-return preferences towards renewable energy projects?: A choice experiment in Germany. *Energy Policy, 97*, 310–320. doi:10.1016/j.enpol.2016.07.042.
Salm, S., & Wüstenhagen, R. (2018). Dream team or strange bedfellows?: Complementarities and differences between incumbent energy companies and institutional investors in Swiss hydropower. *Energy Policy, 121*, 476–487. doi:10.1016/j.enpol.2018.04.046.
Sauré, D., & Vielma, J. P. (2019). Ellipsoidal methods for adaptive choice-based conjoint analysis. *Operations Research, 67*, 315–338. doi:10.1287/opre.2018.1790.
Sawtooth Software. (2014). ACBC technical paper. Sawtooth Software Inc. https://sawtoothsoftware.com/resources/technical-papers/acbc-technical-paper.
Sawtooth Software. (2019a). *Lighthouse Studio Manual: Lighthouse Studio v9.8*. Software for Web Interviewing. Provo, UT.
Sawtooth Software. (2019b). Report on conjoint analysis usage among Sawtooth Software customers. https://www.sawtoothsoftware.com/download/Conjoint_Analysis_Usage_2019.pdf.
Sawtooth Software. (2020). Lighthouse Studio Help. https://sawtoothsoftware.com/help/lighthouse-studio/manual/.
Sawtooth Software Inc. (2023). *Lighthouse Studio: Software for web interviewing and conjoint analysis*. Provo, UT: Sawtooth Software Inc.
Schlereth, C., & Skiera, B. (2017). Two new features in discrete choice experiments to improve willingness-to-pay estimation that result in SDR and SADR: Separated (adaptive) dual response. *Management Science, 63*, 829–842. doi:10.1287/mnsc.2015.2367.
Schomakers, E.-M., Biermann, H., & Ziefle, M. (2021). Users' preferences for smart home automation – investigating aspects of privacy and trust. *Telematics and Informatics, 64*, 101689. doi:10.1016/j.tele.2021.101689.
Schomakers, E.-M., & Ziefle, M. (2022). Privacy vs. security: Trade-offs in the acceptance of smart technologies for aging-in-place. *International Journal of Human–Computer Interaction*, 1–16. doi:10.1080/10447318.2022.2078463.
Schwoerer, T., Little, J., & Hayward, G. D. (2020). Quantifying expert opinion with discrete choice models: Invasive elodea's influence on Alaska salmonids. *Journal of environmental management, 271*, 110924. doi:10.1016/j.jenvman.2020.110924.
Searle, S. R., Speed, F. M., & Milliken, G. A. (1980). Population marginal means in the linear model: An alternative to least squares means. *The American Statistician, 34*, 216. doi:10.2307/2684063.
Shepperson, J., Murray, L. G., Mackinson, S., Bell, E., & Kaiser, M. J. (2016). Use of a choice-based survey approach to characterise fishing behaviour in a scallop fishery. *Environmental Modelling & Software, 86*, 116–130. doi:10.1016/j.envsoft.2016.09.013.
Shin, H.-S., Callow, M., Dadvar, S., Lee, Y.-J., & Farkas, Z. A. (2015). User acceptance and willingness to pay for connected vehicle technologies: Adaptive choice-based conjoint analysis. *Transportation Research Record, 2531*, 54–62. doi:10.3141/2531-07.
Siegel, C. A., Whitman, C. B., Spiegel, B. M. R., Feagan, B., Sands, B., …, et al. (2016). Looking beyond symptoms and disease activity to define disease severity in inflammatory bowel disease: Results of an IOIBD specialist panel. *Journal of Crohn's and Colitis, 10*, 76-78. doi:10.1093/ecco-jcc/jjw019.104.
Silver, S. D. (2018). Multivariate methodology for discriminating market segments in urban commuting. *Public Transport, 10*, 63–89. doi:10.1007/s12469-017-0169-8.

Sipos, P., & Voeth, M. (2015). Performance, motivation and ability – testing a pay-for-performance incentive mechanism for conjoint analysis. In Sawtooth Software (Ed.), *Proceedings of the Sawtooth Software Conference*. Orlando (pp. 143–157). United States: Sawtooth Software Inc.

Sipple, L. R., Racette, C. M., Schiano, A. N., & Drake, M. (2022). Consumer perception of ice cream and frozen desserts in the "better-for-you" category. *Journal of Dairy Science, 105*, 154–169. doi:10.3168/jds.2021-21029.

Smith, S. M., Gewandter, J. S., Kitt, R. A., Markman, J. D., Vaughan, J. A., Cowan, P., et al. (2016). Participant preferences for pharmacologic chronic pain treatment trial characteristics: an ACTTION adaptive choice-based conjoint study. *The Journal of Pain, 17*, 1198–1206. doi:10.1016/j.jpain.2016.07.008.

Sonnier, G., Ainslie, A., & Otter, T. (2007). Heterogeneity distributions of willingness-to-pay in choice models. *Quantitative Marketing and Economics, 5*, 313–331. doi:10.1007/s11129-007-9024-6.

Speed, J., Welk, B., Comiter, C., & Elliott, C. (2022). Determining patient preferences in the treatment of medication-refractory overactive bladder. *Neurourology and urodynamics*. doi:10.1002/nau.25109.

Speight, K. C., Schiano, A. N., Harwood, W. S., & Drake, M. (2019). Consumer insights on prepackaged Cheddar cheese shreds using focus groups, conjoint analysis, and qualitative multivariate analysis. *Journal of Dairy Science, 102*, 6971–6986. doi:10.3168/jds.2018-16209.

Stan Development Team. (2018). Stan user's guide version 2.18. https://mc-stan.org/docs/2_18/stan-users-guide/index.html.

Stan Development Team. (2020a). RStan: The R interface to Stan. R package version 2.21.2. http://mc-stan.org/.

Stan Development Team. (2020b). Stan - Documentation: Version 2.25. https://mc-stan.org/users/documentation/. Accessed 9 November 2020.

Stan Development Team. (2022). shinystan: Interactive visual and numerical diagnostics and posterior analysis for Bayesian models. https://mc-stan.org/users/interfaces/shinystan.

Steiner, M., & Hendus, J. (2012). How consumers' willingness to pay is measured in practice: An empirical analysis of common approaches' relevance. *SSRN Electronic Journal*. doi:10.2139/ssrn.2025618.

Stone, M., Dimofte, C., Strong, D., Pulvers, K., Crespo, N. C., Leas, E. C., et al. (2022). Evaluating US smokers' willingness to pay for different cigarette packaging designs before and after real-world exposure in a randomised trial. *Tobacco control*. doi:10.1136/tobaccocontrol-2021-057071.

Stone, M. D., Dimofte, C. V., Strong, D. R., Villasenor, A., Pulvers, K., Messer, K., et al. (2021). Tool to assess appeal-aversion response to graphic warning labels on cigarette packs among US smokers. *Tobacco control, 30*, 312–319. doi:10.1136/tobaccocontrol-2019-055520.

Storm-Dickerson, T., Das, L., Gabriel, A., Gitlin, M., Farias, J., & Macarios, D. (2018). What drives patient choice: Preferences for approaches to surgical treatments for breast cancer beyond traditional clinical benchmarks. *Plastic and Reconstructive Surgery. Global Open, 6*, 1-8. doi:10.1097/GOX.0000000000001746.

Sulistyawati, I., Dekker, M., Verkerk, R., & Steenbekkers, B. (2020). Consumer preference for dried mango attributes: A conjoint study among Dutch, Chinese, and Indonesian consumers. *Journal of Food Science, 85*, 3527–3535. doi:10.1111/1750-3841.15439.

Toubia, O., Hauser, J., & Garcia, R. (2007). Probabilistic polyhedral methods for adaptive choice-based conjoint analysis: Theory and application. *Marketing Science, 26*, 596–610. doi:10.1287/mksc.1060.0257.

Toubia, O., Hauser, J. R., & Simester, D. I. (2004). Polyhedral methods for adaptive choice-based conjoint analysis. *Journal of Marketing Research, 41*, 116–131. doi:10.1509/jmkr.41.1.116.25082.

Toubia, O., Johnson, E., Evgeniou, T., & Delquié, P. (2013). Dynamic experiments for estimating preferences: An adaptive method of eliciting time and risk parameters. *Management Science, 59*, 613–640. doi:10.1287/mnsc.1120.1570.

Toubia, O., Jong, M. G. de, Stieger, D., & Füller, J. (2012). Measuring consumer preferences using conjoint poker. *Marketing Science, 31*, 138–156. doi:10.1287/mksc.1110.0672.

Toubia, O., Simester, D. I., Hauser, J. R., & Dahan, E. (2003). Fast polyhedral adaptive conjoint estimation. *Marketing Science, 22*, 273–303. doi:10.1287/mksc.22.3.273.17743.

Traets, F., Sanchez, D. G., & Vandebroek, M. (2020). Generating optimal designs for discrete choice experiments in R: The idefix package. *Journal of Statistical Software, 96*, 1–41. doi:10.18637/jss.v096.i03.

Train, K. (2009). *Discrete choice methods with simulation* (2nd ed.). New York: Cambridge University Press.

Vadali, S. (2016). The performance of competitive and lottery incentive schemes vis-a-vis fixed fee incentive schemes in improving conjoint analysis. *International Journal of Management and Marketing Research, 9*(2), 81-100.

van den Borne, B. H. P., van Soest, F. J. S., Reist, M., & Hogeveen, H. (2017). Quantifying preferences of farmers and veterinarians for national animal health programs: the example of bovine mastitis and antimicrobial usage in Switzerland. *Frontiers in Veterinary Science, 4*, 82. doi:10.3389/fvets.2017.00082.

Veitch, J., Ball, K., Rivera, E., Loh, V., Deforche, B., Best, K., et al. (2022). What entices older adults to parks? Identification of park features that encourage park visitation, physical activity, and social interaction. *Landscape and Urban Planning, 217*, 104254. doi:10.1016/j.landurbplan.2021.104254.

Veitch, J., Ball, K., Rivera, E., Loh, V., Deforche, B., & Timperio, A. (2021). Understanding children's preference for park features that encourage physical activity: an adaptive choice based conjoint analysis. *The international journal of behavioral nutrition and physical activity, 18*, 133. doi:10.1186/s12966-021-01203-x.

Veitch, J., Salmon, J., Deforche, B., Ghekiere, A., van Cauwenberg, J., Bangay, S., et al. (2017). Park attributes that encourage park visitation among adolescents: A conjoint analysis. *Landscape and Urban Planning, 161*, 52–58. doi:10.1016/j.landurbplan.2016.12.004.

Wertenbroch, K., & Skiera, B. (2002). Measuring consumers' willingness to pay at the point of purchase. *Journal of Marketing Research, 39*, 228–241. doi:10.1509/jmkr.39.2.228.19086.

Wicki, M., & Kaufmann, D. (2022). Accepting and resisting densification: The importance of project-related factors and the contextualizing role of neighbourhoods. *Landscape and Urban Planning, 220*, 104350. doi:10.1016/j.landurbplan.2021.104350.

Will, C., Lehmann, N., Baumgartner, N., Feurer, S., Jochem, P., & Fichtner, W. (2022). Consumer understanding and evaluation of carbon-neutral electric vehicle charging services. *Applied Energy, 313*, 118799. doi:10.1016/j.apenergy.2022.118799.

Wuebker, R., Hampl, N., & Wüstenhagen, R. (2015). The strength of strong ties in an emerging industry: Experimental evidence of the effects of status hierarchies and personal ties in venture capitalist decision making. *Strategic Entrepreneurship Journal, 9*, 167–187. doi:10.1002/sej.1188.

Yang, L., Toubia, O., & Jong, M. G. de. (2018). Attention, information processing, and choice in incentive-aligned choice experiments. *Journal of Marketing Research, 55*, 783–800. doi:10.1177/0022243718817004.

Yu, J., Goos, P., & Vandebroek, M. (2009). Individually adapted sequential Bayesian designs for conjoint choice experiments. *SSRN Electronic Journal*(April), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1395663.

Yu, J., Goos, P., & Vandebroek, M. (2011). Individually adapted sequential Bayesian conjoint-choice designs in the presence of consumer heterogeneity. *International Journal of Research in Marketing, 28*, 378–388. doi:10.1016/j.ijresmar.2011.06.002.

Zenker, S., Eggers, F., & Farsky, M. (2013). Putting a price tag on cities: Insights into the competitive environment of places. *Cities, 30*, 133–139. doi:10.1016/j.cities.2012.02.002.