

THE DIFFERENCE BETWEEN 'LESS BAD' AND 'MUCH BETTER'

HELPING CONJOINT TO LIVE UP TO ITS PROMISES BY LEVERAGING 'BEHAVIOURAL ECONOMICS'

Florian Bauer

PREFACE

Recently, market researchers have become increasingly interested in integrating insights from 'Behavioural Economics' into their traditional research approaches. However, this was pretty much restricted to an inductive transfer of selected insights into designs and questionnaires on a case-by-case basis.

Moreover, there was no systematic merging of these insights with the classical blockbuster tools of market research. Conjoint analysis is definitely such a blockbuster tool and probably the one with the most advanced analytical basis. What we will show here is how conjoint analysis can for the first time be seamlessly merged with the most important insights from 'Behavioural Economics', making predictions more valid while maintaining the core advantages of conjoint analysis. We will argue that instead of continuously searching for better conjoint variants that are 'less bad' than previous ones, 'much better' solutions can be found by merging the strengths of these divergent or, at the very least, complementary approaches.

INTRODUCTION

In the days when conjoint analysis was still called conjoint measurement, its primary aim was to verify if human choices met the conditions of rational choice. They did not. Ironically, since then conjoint analysis has had an unmatched career in marketing research as the go-to tool to simulate and model human choice, despite the fact that all existing versions of conjoint still assume rational choice –something that the original conjoint measurement version proved to be wrong!

While 'Behavioural Economics' and conjoint analysis both have the same roots – a desire to better understand human choices – their development since those early days could not be more different: conjoint analysis simulates human choices under the assumptions of rational choice (stable and absolute preferences, perfect information, etc.) whereas 'Behavioural Economics' has developed by collecting more and more insights about where, when and why people are making predictably irrational decisions.

Today, conjoint analysis is undoubtedly one of the most important and most advanced tools in market research. This indispensable tool has witnessed several significant improvements in recent years but the results it produces are still compromised by two inherent distortions: 'inherent' because they are hard wired into the conjoint-specific data-gathering procedure; 'distortions' because these procedures are incapable of sensitively capturing the sub-optimal decision effects that have been proven by countless experiments in 'Behavioural Economics'.

PROBLEM 1: COGNITION-RELATED DISTORTION

The first inherent distortion is cognition-related and concerns the fact that in the real world the subjective utility of quantitative attributes (such as the price of a given product, a car's bhp, or a hard drive's storage capacity) typically follows a step-like pattern: the subjective value changes abruptly when specific threshold values are reached, which is why a price of $ 0.99 appears to us to be much more attractive than a price of $ 1, despite the difference being only 1 cent. In 'Behavioural Economics' this is known as a typical 'assimilation and contrast effect'. Nevertheless, conjoint analysis completely neglects effects like this. Instead, it reconstructs subjective utility by testing discrete attribute levels between which utility values are then interpolated in a linear fashion. This linear interpolation means that the 'stepped' utility function unavoidably disappears. Clearly this leads to critical misinterpretations because behaviourally relevant thresholds and indifference zones are neglected.

Example: If a conjoint analysis revealed that 80% of customers would buy a certain CD for $ 9, but only 50% at $ 10, it would then, by means of linear interpolation between the two tested attribute levels, conclude that only 53% would buy it at $ 9.90. However, pricing experience as well as tools like Peter van Westendorp's classical 'Price Sensitivity Meters' (PSM) show that there is a high chance that there will be a significant price threshold at $ 9.90, making it a much bigger difference between $ 9.90 and $ 10 than linear interpolation would imply.

PROBLEM 2: MOTIVATIONAL DISTORTION

The second inherent distortion typical of conjoint analysis is motivation-related. It concerns the fact that it is impossible within the context of conjoint analysis to capture effects of differing product or decision involvement in a valid way: In conjoint analysis the structured presentation of options, with all attributes being clearly and comparably described, creates an artificial decision context inducing significantly more involvement than most people would have in a real-life context, where they often do not spend much time in scrutinizing and comparing all potentially relevant attributes and options. In contrast to that, within a conjoint analysis, respondents find themselves in a situation that artificially evens out difference in involvement. From the perspective of 'Behavioural Economics' this is known as a typical 'framing effect'. Even if some differences due to differing levels of involvement 'survive' this artificial framing of choices, they are finally wiped out arithmetically by the individual standardisation of partial utility values. Hence, the partial utility sensitivities of respondents do not reflect real-life preferences, as involvement effects are minimized by the typical framing of choices in the conjoint sequence and the mathematical neutralization in the subsequent analysis and modelling.

Example: We design a conjoint analysis and let a specific respondent – let's call him 'Mr. Miller' – chose between a free bank account offered by Bank A (with no interest paid if he's in credit) or the account proposed by Bank B, with annual charges of $ 19.90 but 1% interest paid on accounts.

When presented with this choice set as part of the conjoint sequence, Mr. Miller possibly feels he's been transported back to his maths class. He works out that account B makes sense for him if his average balance is above $ 1,990, and on this basis he thus puts a cross next to this option. In the end, this had very little to do with a real-life decision situation. In real life, Mr. Miller has never given a second thought to what his bank account costs. He's had an account at his local bank for the last 17 years, with no interest paid and annual charges of just under $ 120 ($ 9.90 a month).

So to begin with, Mr. Miller in real life is not aware of the charges associated with his account (he has no price knowledge), and secondly he would never dream of running from one bank to another, comparing the terms and conditions of various bank accounts (he has no price interest). If this lack of price knowledge and price interest applies not only to Mr. Miller, but to other customers or even the majority of customers, conjoint analysis no longer produces realistic results because of the involvement that's induced – even in people like Mr Miller once they are forced to run through a conjoint sequence. This is a classical effect (or rather bias) we can find in countless conjoint analyses.

SOLUTION 1: 'COGNITIVE PATCH'

The 'General Algorithm for Patching Conjoint Analyses' (GAP), corrects the cognition-related distortion with the help of the so-called 'Value Sensitivity Meter' (VSM), a generalised variant of the 'Price Sensitivity Meter' (PSM). The advantage of VSM compared to conjoint analysis is the unprompted approach: minimum and maximum values are identified through open-ended questions. In generalising PSM, VSM can be used for all quantitative attributes and not only price, i.e. for all attributes where the subjective utility function is actually rather step-like than linear. VSM accurately identifies thresholds and indifference areas by not only asking whether a product is too expensive, expensive, cheap or too cheap like in PSM but by adapting the questions to whatever quantitative attribute is appropriate, thus for example, (too) big/small, (too) many/few, (too) fast/slow, etc. Moreover, only two of the typical four questions are necessary, as the thresholds which emerge tend to be almost identical for all four questions. Therefore, two questions are sufficient to cross validate thresholds and indifference areas.

Subsequently, the advantages of this approach are illustrated in an example analysing the optimal horsepower for a car – a simplified case study directly taken from a real project: Participants were asked how much horse power they consider to be too weak and too strong. Experience showed that the values on the y-axis (left side of figure 1) cannot be taken seriously but those on the x-axis should: the insight gained from VSM is not the percentage of agreement or disagreement but the position of the attribute thresholds and indifference areas as illustrated on the right side of figure 1.

The locations of thresholds and indifference areas are exactly the pieces of information which are not given by conjoint analysis alone. Thresholds and indifference areas from VSM, once "projected" onto the x-axis, can mathematically be combined with the partial utility values of distinctly tested attribute levels within conjoint analysis. Graphically speaking thresholds and indifference areas replace the linear interpolation typical to conjoint analysis (see figure 2).

FIGURE 1, INVALID AND VALID INSIGHTS FROM VALUE SENSITIVITY METER (VSM)

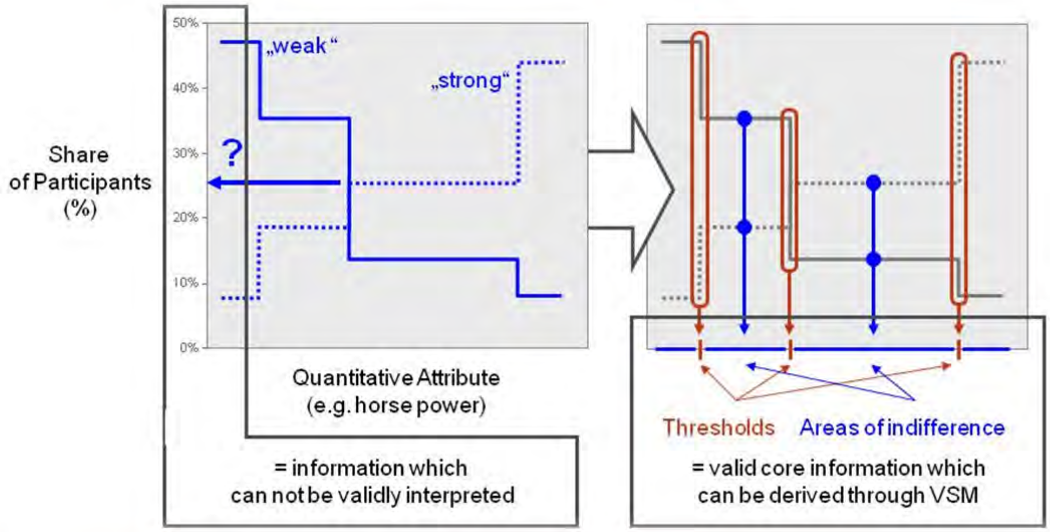
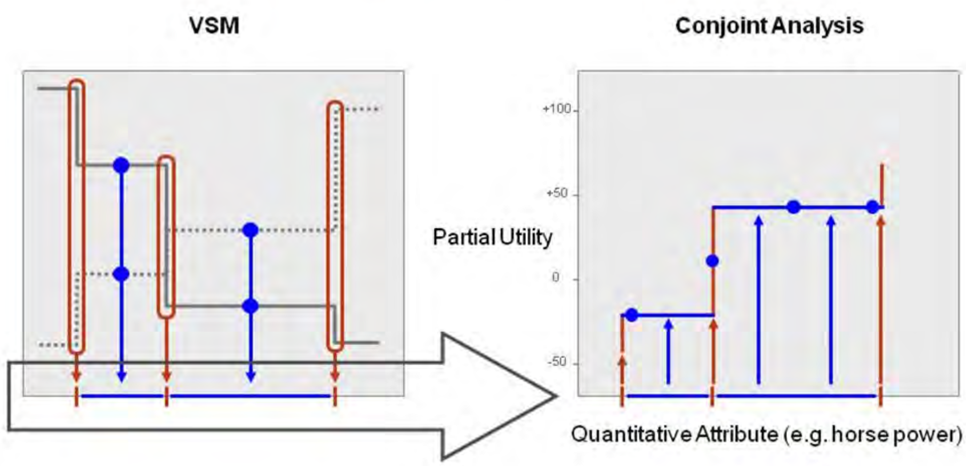


FIGURE 2, COMBINING CONJOINT ANALYSIS AND VALUE SENSITIVITY METER (VSM)

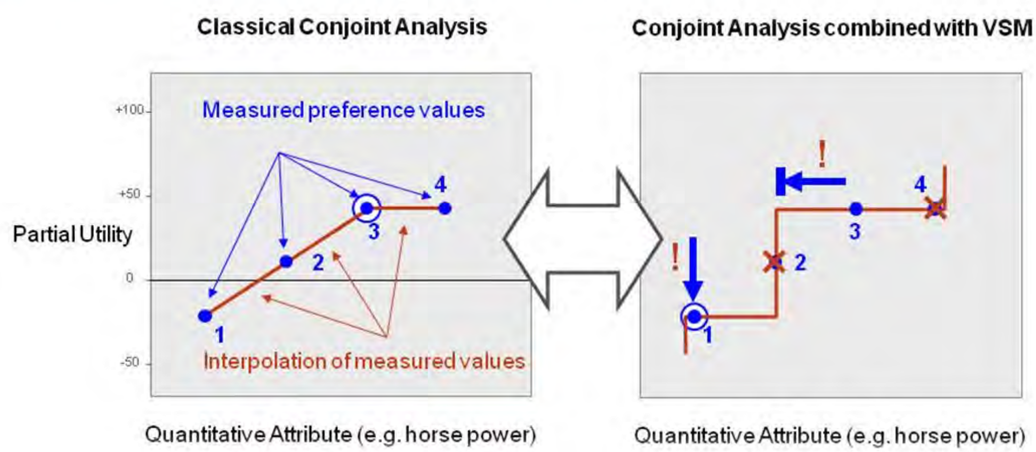


Merging both insights efficiently leverages the complementary strengths of VSM and conjoint analysis while simultaneously neutralizing their weaknesses: all simulations typical for conjoint simulations can be done in exactly the same way as before (even with standard conjoint analysis software), but they are now based on modified utility functions that reflect thresholds and indifference areas that would have otherwise been overlooked. Consequently, recommendations which are derived from this seamlessly combined analysis could not have been revealed by classical conjoint data alone (see figure 3).

In fact, using conjoint data alone would even have led to the wrong recommendations being made in our example of determining the optimal horsepower for a new car (see right side of figure 3):

- In the combined analysis, attribute level 1 is a local optimum
- Level 2 is suboptimal, as it is right on a threshold in this case. Instead, only slightly right of level 2 would be another local optimum as it would be just beyond the identified threshold.
- Level 3 again is not optimal: high horse power, thus more production costs, with no effect on the utility level.
- Also level 4 should not be recommended but slightly increased to move it just right of the threshold. A minimal increase of horsepower would have a major effect in partial utility.

FIGURE 3, TWO METHODS SEAMLESSLY COMBINED FOR NEW INSIGHTS



None of these recommendations could have been derived out of the linear interpolation of the conjoint analysis alone. In a linear interpolation, as conducted in a conjoint approach, the sensitivity curve would have increased continuously from attribute level 1 to 3, and would have then continued horizontally to attribute level 4. Most interestingly, conjoint alone would have suggested that level 3 would have been the local optimum! Only the 'cognitive patch' within GAP shows that the truth lies in between.

This simple correction, which is achieved by merely adding a few VSM questions before the actual conjoint sequence and by conducting a slightly more complex analysis, is duly justified, as it makes more valid findings possible, as the human factor of 'cognitive distortion' is factored into the results.

SOLUTION 2: 'MOTIVATIONAL PATCH'

Involvement-related sensitivity differences that are evened out or ignored by conjoint analysis can be quantified equally simply. The basic requirement is an experimental, monadic design 'on top of' the conjoint analysis which is realised in the frame questionnaire 'outside' of the actual conjoint sequence. This allows us to identify preference sensitivities between pre-defined product options ('hold outs') that are influenced by actual differences in involvement, independently: in its simplest form, each respondent indicates his or her willingness to buy a specific product before the actual conjoint sequence starts. Additionally, the involvement of each participant is measured on a standardized scale. Table 1 shows the simple resulting 2-by-2-design if we use two 'hold outs' across respondents (attractive versus unattractive options) and dichotomize involvement (high versus low involvement).

TABLE 1, PURCHASE INTENTION DEPENDENT ON PRODUCT ATTRACTIVENESS AND INVOLVEMENT IN A SIMPLIFIED 2-BY-2-DESIGN

Purchase Intention (PI)	High Involvement (HI)	Low Involvement (LI)
Product Option 1 (‚attractive')	PI_{1HI}	PI_{1NI}
	Sensitivity HI	Sensitivity LI
Product Option 2 (‚unattractive')	PI_{2HI}	PI_{2NI}

The core notion behind this approach is that extremely low involvement will lead to basically random choice and hence minimal sensitivities between product options with differing attractiveness. Hence, through comparison of the purchase intention relation between high versus low involvement participants, and thus through PI_{1HI}/PI_{2HI} versus PI_{1NI}/PI_{2NI}, the sensitivity reducing influence of low involvement can be directly quantified on the basis of this monadic experimental design. For example, actual influence of (low) involvement is simply measured by comparing the revealed preference sensitivities for the differing product options of highly involved respondents versus low involvement respondents. The typical result is that participants with low involvement show significantly lower sensitivities than those with higher involvement. These sensitivity differences are 'lost' in the partial utilities coming out of the classical conjoint analysis and need to be 're-projected' into the simulation of preferences. This is done by moderating the probability parameter so that the option with the highest theoretical utility is actually preferred. In the case of high involvement, for example, the probability of an option with theoretically higher utility to be actually chosen tends to be higher as respondents with higher involvement pay more attention to attribute and product differences. This probability parameter within the conjoint simulation is hence moderated in such a way that we reproduce the same preference sensitivities within the conjoint simulation as we quantified in the experimental design outside of conjoint – for each involvement segment that was differentiated in the experimental design, respectively (two segments in our simple example: high versus low involvement). Finally, the results of the segment-specific simulations only have to be merged again in an appropriately weighted way.

A huge advantage of this approach is that the impact of involvement is not predefined but exactly quantified in each project, logically guaranteeing maximum validity as each patch is basically validating itself every time: I.e. if involvement differences do not moderate preferences, there will be no moderation based on this approach.

While the 'cognitive patch' moderates the partial utility function, the 'motivational patch' technically moderates the conversion of utilities into preference scores within the conjoint analysis. In the case of both patches, the insights gained outside of the actual conjoint analysis are directly projected into the conjoint analysis, allowing the same simulations to be used but with more valid input data.

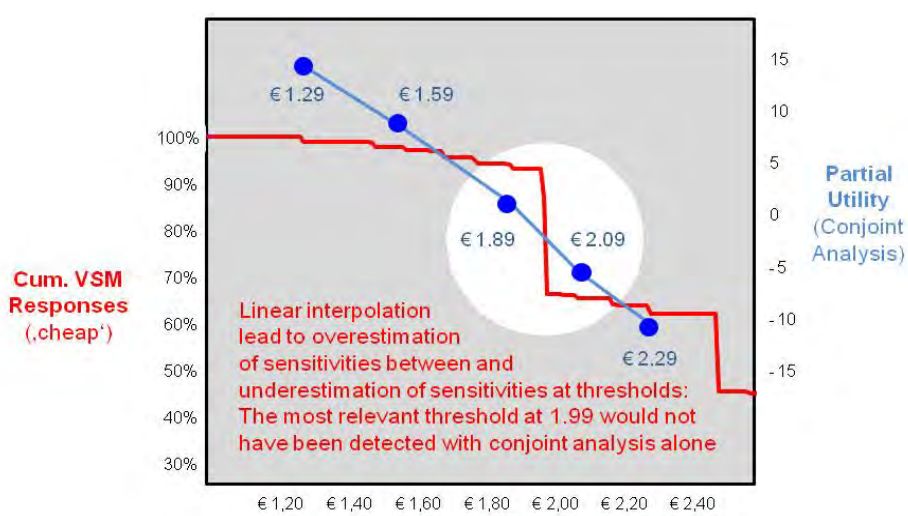
VALIDATION

However, one crucial question remains: Does 'GAP' increase the validity probability compared to a traditional conjoint analysis?

In fact, this question can be answered in the positive without even considering any empirical data. To see why, let's first take a closer look at the 'cognitive patch': The cognitive patch is not changing actually measured results, i.e. it is not interfering with the results conjoint delivers for the actual tested attribute levels. GAP is only exchanging an often falsified conjoint assumption (i.e. 'utility follows a linear function between two attribute levels') with actual data from a classical and frequently validated method to measure utility thresholds (i.e. VSM/PSM). Hence GAP brings in more empirical data in exchange for an assumption that is equally frequently proven to be wrong as a threshold-function is proven to better map individual utility functions. Seen from that perspective, the burden of proof rather lies on conjoint analysis than on GAP.

Despite this, we collected a lot of validating data from conjoint on FMCG: a research area where conjoint analysis is known to deliver notoriously biased results but is nevertheless often used because of its multidimensional modeling options. In this case, we simulated the effects of different price points for a multipack of candy bars. We tested all relevant attribute levels from the client's perspective and also ran a classical VSM in parallel. The results are depicted in figure 4. Conjoint did not indicate any thresholds due to the linear interpolation between the tested attribute levels, while VSM clearly identified the threshold at 1.99 for example, which eventually turned out to be the (by conjoint alone, unidentified) optimum. The thresholds of VSM were perfectly in line with subsequent sales data but the 'take rate' was greatly overestimated by VSM. The opposite held true for the conjoint data: The 'take rate' or market shares predicted for the actually measured price points, were good, but everything in between was incorrect (due to the insensitivity to thresholds). Having the strengths of both, conjoint and VSM, combined in GAP produced the best fit and predicted critical price thresholds that prevented the client from falling into this classical conjoint trap.

FIGURE 4, VALIDATING GAP ('COGNITIVE PATCH'



We often find that the optimal attribute level is identified as lying between two actually tested attribute levels and that, if this is the case, the 'cognitive patch' will frequently lead to a different recommendation than that suggested by classical linear interpolation. We have seen this happen in many different projects and industries ranging from automotive to FMCG.

Similarly, there is logical proof and empirical evidence for the significantly higher validity of the 'motivational patch' which is the second part of GAP. In fact, there are two logical arguments: Firstly, what the 'motivational patch' does is theoretically the same as what every classical conjoint analysis does to calibrate preference share into predicted take rates: This calibration is always done by using the same kind of hold-out tasks as the 'motivational patch'. The only thing the 'motivational patch' is doing differently is that it follows a more complex design that allows for additional differentiation between different involvement segments. Hence, we measure something that has previously not been paid attention to, then we integrate it in the same way as similar calibrations have been proven to be integrated best in classical conjoint analysis.

Secondly, the core advantage of this patch is that it is sensitive to the given situation, i.e. how far the conjoint results are patched is effectively dependent on the actual choice sensitivity of respondents with differing involvement levels. If respondents with low involvement happen to show the same choice sensitivity in the hold-out tasks as people with high involvement, then the simulation will not be moderated. There is no fixed parameter that always defines the necessary moderation and that would need to be empirically validated.

So far we have made the logical arguments for the 'motivational patch', but now let's take a closer look at some real data: In a very typical conjoint analysis on mobile telecommunication tariffs we differentiated between high and low involvement customers. We firstly had a look at the uncorrected conjoint results: Looking at importance levels (and the underlying sensitivities) there were absolutely no significant differences (see figure 5). But when we compared the results of the hold-out tasks, it turned out that the high involvement customers were very sensitive to the differences of the monadically tested options while the low involvement customers' preferences did not differ between the objectively divergent options if they could not be directly compared. Now, the question is which result is more valid? In order to answer this, we always ask some additional questions in the frame questionnaire. For example, here we questioned the customer's actual market price knowledge as an external parameter: It turned out that although all of them were recruited as customers who were currently in the middle of an actual decision process to do with their mobile telecommunication contract, the difference in price knowledge between the high and the low involved customers was extremely wide. While the high involvement customers had a quite decent price knowledge and were able to judge whether any given offer was rather high or low priced, the low involvement customers turned out to be unable to judge this reliably even though they were all about to make a purchase decision for themselves in 'real life'.

FIGURE 5, VALIDATING GAP ('MOTIVATIONAL PATCH')

Parameters	High Involvement (HI)		Low Involvement (LI)
Conjoint Attributes			
Recurring charges	19.1		19.0
Brand	14.9		14.2
VAS	10.9		11.8
Attribute 4	12.0		10.9
Attribute 5	11.0	=	10.2
Attribute 6	9.2		9.6
Attribute 7	6.7		7.8
Price per Minute	8.9		8.8
Handset	7.5		7.7
'Emotional Patch':		Contradiction	
Willingness to buy Hold-out Task 1 ('attractive') (1= high; 6=low)	2,1		2,5
		≠	
Willingness to buy Hold-out Task 2 ('unattractive') (1=high; 6=low)	3,5		2,7
		Validation	
External Parameter: Price knowledge 1=good; 6=bad)	1,9	≠	4,1

This firstly explained the differing results of conjoint versus hold-out: Conjoint analysis was inducing perfect price knowledge and so the choice task within conjoint was merely a comparison task where high and low involved customers could perform equally well, while in the 'hold-out' tasks their actual expertise led to very different choices. Secondly, the price knowledge data analysed in parallel, validated that the results of monadic hold-out tasks are much more realistic than conjoint which is significantly biasing the choice situation by inducing perfect price knowledge even in customers that do not have such a level of expertise in reality.

A final 'proof' of the validity of the results of GAP is that none of our clients has ever disregarded the results of the patches: none would rather just follow the uncorrected conjoint results. Alone, this might not be a sufficient proof, but for us it is a crucial aspect, and together with the logical and empirical validation, shows that GAP can be an indispensable improvement of conjoint analysis.

CONCLUSION

The 'General Algorithm for Patching Conjoint Analyses' GAP efficiently corrects, or patches, the main cognitive and motivational distortions which occur in conjoint analysis. This allows significantly more valid predictions to be made, as the gap between the idealized decision situation of conjoint and the reality of the actual purchasing behaviour is closed. GAP is unique and promising as it seamlessly bridges the gap between two complementary worlds: excellence in research and modeling and excellence in understanding human decision making. The strengths of the 'rationalistic' conjoint approach are seamlessly combined with insights from 'Behavioral Economics' on how cognition and motivation make people deviate from what is considered a rational decision.

GAP is not merely an extremely efficient procedure, it is also 'open source' and universally valid insofar as it is not a 'black-box' approach but one everybody can employ right away as it can be combined with any conjoint variant.

Finally, this approach opens up a new and complementary direction for future developments: Until now conjoint improvements and enhancements were driven by the attempt to develop even better conjoint analysis. This indeed resulted in even more precise variants (such as ACBC, HILCA, ISBC, etc.). However, all variants still are, and will forever be, subject to the mentioned cognitive and motivational distortions, as they are methodically inherent in the approach. Thus a real improvement is only possible by combining the strengths of alternative methods with the advantages of conjoint analysis itself. Instead of continuously searching for better conjoint variants that are 'less bad' than the previous ones, this demonstrates how 'much better' solutions can be found (with less effort) if one finds a smart way of merging the results of divergent approaches.

THE AUTHOR

Florian Bauer, Member of the Board, Vocatus AG, Germany, and Member of the Board of the German Market Research Association (BVM).