

Report 1, ‘Theoretical markers of harm for machine play in a bookmaker’s – a rapid scoping review,’ formed the initial academic and theoretical basis for the Responsible Gambling Trust’s B2 Machines Research Programme. Of all the document provided as part of the programme, Report 1 is the most cautious in outlining the limitations of the approach decided on. It is the least politicised, as well as being articulate about the gaps in knowledge and the need for continued thought to be given to the approach.

Within two weeks, researchers sought to construct a framework to analyse an enormous volume of data in such a way that two questions posed by the RGSB could be answered. Unfortunately, it is clear from this report that answering the second question was never going to be a possibility, due to the metrics chosen.

The background for this is curious, as is evident in the analytical commentary. It appears that certain methodological choices were made to enable such a vast research programme to commence in an ordered, scientific way – but that some of those choices were less than optimum.

There was a clear opportunity for several steps to be made toward answering the second question posed by the RGSB and investigate what measures could be taken to prevent harm. Despite the need for time considerations, including a brief analysis of debit transactions and incorporating more surveys and contextual information into the final programme would have given future researchers a leg up in addressing harm-prevention measures. Repeated debit transactions are seen as a plausible risk, and contextual metrics are crucial in assessing ‘gambling harm’ rather than ‘problem gambling’ – both of which distinctions are evident in Report 1. However, subsequent analysis dismissed using these as variables for consideration.

In addition, between-session markers of harm were shown to be better predictors of ‘problem gambling’ than within-session metrics by the researchers’ own criteria scheme. However, all session analysis for non-carded play (i.e. the majority of that undertaken in Report 3) were examined via the computation of a ‘proxy session’. The limitations of this approach are discussed in Report 3, but in summary, it means that the methodological weaknesses are compounded throughout the research programme.

This is a noted methodological weakness of the program. The within-session markers of harm were known to be supported by less and weaker evidence than between-session markers. Moreover, sessions for non-carded play could only be determined by proxy (see limitations to proxy sessions, Report 3). Finally, the majority of the events received were not linked to carded play: therefore, the bulk of the research was based on weak metrics analysed via a suspect proxy session calculation.

Another area of concern is the way all metrics were evaluated to determine if including them in the research was appropriate. The Bradford Hill criteria was used to provide a baseline against which metrics could be measured for strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment, and analogy – criteria developed to assess causation in clinical trials. Using these criteria to assess the suitability of metrics for a sociological framework undoubtedly raises questions about the sustainability of that framework.

Nonetheless, despite these choices Report 1 presents the most balanced assessment of what may and may not be possible to determine in gambling data, and the limitations inherent in undertaking such a programme. It strongly supports the need for a contextual understanding of all interpretations of markers or patterns of harm, but fails to lead into a research framework for just such an understanding throughout subsequent analyses – a great loss not just to the research programme commissioned, but for future studies on harm minimisation and prevention as well.

Report 2, 'Identifying problem gambling – findings from a survey of loyalty card customers',

presented the researchers involved in the RGT-commissioned programme with a rare opportunity to investigate the responses of loyalty card holders. The survey included questions which measure the extent of problems experienced by those who gamble. The writers of the document report a number of aims:

- To document the survey process
- Give an overview of the broader gambling behaviour of loyalty card holders
- Identify the prevalence of problem gamblers among loyalty card holders
- Introduce key themes used in predictive analysis (i.e., signpost the approach taken in Report 3 for audiences)
- Highlight caveats of the research

The survey process involved randomly sampling 27,565 loyalty card holders and requesting survey participants, of which 4,727 individuals agreed. Survey responses were then linked with loyalty card data for 4,001 of those individuals, for a response rate of 17-19%. Although the innovation of the RGT's research programme was consistently trumpeted, this is relatively standard practice.

The broader gambling behaviour of loyalty card holders provided paints a troubling picture, especially given that this group formed a critical part of the analysis conducted in Report 3. The researchers found that 21% of those surveyed had more than one loyalty card and that 68% of all individuals surveyed didn't always use their card while visiting and LBO. Any conclusions drawn from this group must necessarily be shadowed by confusion, as it is impossible to tell from gambling data whether or not a single individual is actually in possession of the loyalty card being used. Additionally, the subsequent analysis conducted by researchers produced strong indications that some loyalty cards were being used to launder money.

As is typical of this particular research programme, definitions and conclusions were carefully crafted in such a way as to advocate for no immediate industry action to be taken and to call for further research. Defining problem gambling as "gambling to a degree that compromises, disrupts or damages family, personal or recreational pursuits" does not encompass at-risk gamblers whose behaviour is indicative of an impending difficulty. Indeed, although the PGSI screen used to conduct part of the survey is meant to assess problem gambling along a spectrum of harm, Report 2 focuses on sustaining a dual concept of gambling addiction: problem vs. non-problem gambling. This may in part be due to the questions originally proposed by the RGSB as the basis for the RGT's research programme, but also indicates an overall determination (evident in all seven reports) to bifurcate gambling addiction into simplistic terms that refocus the onus of prevention on the individual. (In other words, if there are only two categories – problem and non-problem gamblers – then those who experience problem gambling simply need to adopt the behaviours of non-problem gamblers.) This is miles away from an interpretation of available data that would form the basis for 'corporate social responsibility' and instead supports the gambling industry status quo of focusing on individual behaviour.¹

This selective, narrow approach is never more apparent than in the below paragraph (page 12).

¹ This also supports Featurespace's approach, popular in the gaming industry for its fraud prevention effectiveness, which takes individual behaviour as a basis for analysis. As Featurespace was a key partner in this research, no consideration was given to the notion that a non-individualised approach (such as one that included community as well as corporate intervention) may be more suited to effective harm intervention. It is equally interesting to consider that Featurespace's approach can therefore be considered the antithesis of harm prevention, and is instead fully occupied with identifying 'problem gambling' after it has occurred.

“To date, regulation of gambling machines tends to be conducted at a fairly blunt level and focuses on restrictions of stake, price, speed, and numbers of machines in certain venues. There is no regulation that is tailored to individual gamblers. The Gambling Commission (the industry regulator) considers that mix of macro (e.g., stakes and prizes) and micro (e.g., the individual) regulatory approaches may be effective. Therefore, a critical question is whether industry data can identify ‘harmful’ patterns of play at an individual level and if so, what types of interventions could be introduced that intercede with gamblers experiencing problems. A further concern is to ensure that any individual-led policies intervene with those experiencing problems, whilst allowing those who are not experiencing problems to gamble without onerous intervention.”

This statement can be parsed to indicate the biases inherent in the programme. By indicating that regulations to date have been blunt measures focusing on one metric, the researchers lend support to their thesis that a combination of factors must be considered. Indeed, acknowledging that the Gambling Commission favours macro and micro measures would seem to support this. However, there is no indication that the opening sentences in this paragraph logically lead to there being a critical question around whether harmful patterns are identifiable at an individual level. Indeed, investigation whether stakes and prizes indicate harmful patterns would, therefore, be equally important (according to the Gambling Commission). Nonetheless, the focus continues to be on intervening at an individual level while not labouring non-problem gamblers with ‘onerous intervention.’

Given NatCen’s self-description as ‘an independent, not for profit organisation [able] to put our time and energy into delivering social research that works for society,’ the question therefore arises: what is so horrible about preventing problem gambling?

If the ‘thresholds’ and ‘trade-offs’ discussion which began in Report 1 and is fully expounded in Report 3 is to be read critically, the researchers indicate that it is an established part of the programme to identify problem gambling only in those who are currently experiencing harm. To send a responsible gambling message to a ‘non-problem gambler’ (and again, there are difficulties in this stark duality recognised in this very report) may serve a preventative purpose and increase the overall health of the gambling population. As researchers argue that “Gambling operators should think carefully about the level and type of promotions offered to [loyalty card] customers” (page 11), it is evident that a message is being delivered. To wit: with the right kind of interventions, you can address only those gamblers whose problems are so extreme as to be beyond ignoring. By not intervening until that extreme moment, at-risk customers can be encouraged to contribute to operator revenue for as long as possible.

Operators are also encouraged to collect contextual information about their customers to improve the identification of problem gamblers. Loyalty card customers already represent a proportion of the population at an ‘elevated risk of experiencing problems from gambling’ and therefore allowing operators access to more data provides further avenues for contact, including increased opportunities for sending communications which may exacerbate gambling harm. This is very hard to swallow from an organisation touting their independence and commitment to research that ‘[makes] life better.’

Report 2 also echoes the limitations felt by the researchers as regards time and scope, citing the 5-week turnaround between commissioning and surveys as leaving no opportunity to develop questions aimed at measuring gambling-related harm. However, the researchers consistently ignore previous work that was conducted outside of the UK, including surveys that would give insight into the more contextual areas of harm identification and prevention. Rather, the researchers claim that

‘there has been little work aimed at quantifying and measuring this broader range of gambling harms and there are no validated survey questions which can be used’ (page 15); this claim allows them to propose the Problem Gambling Severity Index (PGSI) as a measurement of problem and at-risk gambling that all stakeholders could agree on. Maintaining the thin veneer of transparency, the researchers admit that using this screen alters the aims of the research. What is not so apparent is that these subtle changes (in definitions, in metrics chosen, in measuring ‘problem gambling’ vs ‘gambling harm,’ and so on) nudge the reader over the course of all seven documents toward believing that there is already a viable solution in place for combatting gambling addiction: Featurespace’s analytical approach.²

Bridging the shaky theoretical ground of Report 1 and the blatantly commercial bias of Report 3, Report 2 delivers troubling results and attempts to box them up in a confusion of noise. Response rates are reported to be low, while the research is touted to be the largest of its kind. Loyalty card holders are described as ‘highly engaged’ (meaning that analysis of their play may not reflect the general population of players), while at the same time there are questions around how ‘representative’ survey responses are for loyalty card holders.

The ‘number of gambling activities’ examined is a construction that is rife with disappointment. Rather than focus on machine play, attention is diverted to the range of various gaming activities that ‘heavily engaged’ and ‘at risk’ loyalty card holders may participate in. Gambling activity specifically undertaken on bookmakers’ machines is compared to participation in ‘most frequent activity’ (page 35), which begs the question: what special attention will be devoted to those whose most frequent activity *is* gambling on bookmakers’ machines?

Although it isn’t specifically highlighted in the report, those groups for which the most frequent activity *is* FORT gambling are arbitrarily given the distinctions of Class 3 (‘Substantial engagement gamblers’) and Class 4 (‘Heaviest engagement gamblers’). Class 3’s favourite form of gambling is machines in bookmakers, followed by at least five other types of gambling preferences. 97% of Class 4 individuals gambled on FORT machines, followed by as many as nine other activities. Despite the researcher’s attempt to contextualise ‘substantial’ and ‘heaviest’ engagement gamblers among the number and variety of other types of gambling undertaken, these assessments are the closest the research comes to admitting that the product is problematic (page 41-42).

The researchers recognise the limitations in using surveys where human interviewers administer questions, a known caveat in social science research that encourages respondents to give answers which they think are ‘socially desirable’ (and therefore not indicative of actual levels of play, engagement, expenditure, harm, etc.) especially where addiction is concerned. Nonetheless, they chose to implement phone surveys. Disappointingly, the phone surveys conducted involved the PGSI, which has not been validated in Great Britain and has implications for the accuracy of identification.

² A final word about the ‘unique contributions’ of Report 2, as mentioned both within the body of the report and the summary provided to the RGT for its December 2014 conference. Although the volume and variety of data analysed is ground-breaking for Great Britain, much of the methodology relied on outdated measures of those that were the most familiar to the research team at the time of project design (see commentary on Bradford Hill criteria, Report 1 and Report 1 Summary). The lead author on Report 2, Heather Wardle, was also the project director for the British Gambling Prevalence Survey and advocated for a conservative approach to framework that would exclude newer (post-2010) and international (outside of Great Britain) research and survey methods.

Report 3 ‘Predicting Problem Gamblers; Analysis of industry data’ as part of the RRG T B2 Machines Research Programme focuses primarily on the initial results obtained by Featurespace from an analysis of data provided by industry operators and gaming machine manufacturers. As indicated in the comments, this analysis was conducted at the very end of the six-month programme, which resulted in a number of unanswered questions and suggestions for further research. It cannot be known what additional insights would have been gained had more time been allotted.

When this document was first published, PR concern did not centre on whether or not the research itself could be criticised, as researchers were confident that the report was without defect. Therefore, the focus was on whether or not the researchers themselves would be attacked: for example, the long-term association between the RGT and Featurespace, or Featurespace’s historical involvement with Betfair.

However, in reviewing the comments provided in the analysis it should be apparent that there may be some basis for critique of the research itself, especially when taken as one single piece of a research programme designed to support the status quo and not ruffle the feathers of industry organisations (i.e. potential future customers). Therefore a decision was made to ‘throw the hat in the ring’ with the industry and frame the entire document with a political agenda, making it clear to any audience that the £2 stake limit proposed by lobbying groups and local councils would be quashed by ‘independent, scientifically-backed research’. This was certainly effective, as Newham Council’s campaign was rejected in mid-July of this year.

Before detailing the methodological difficulties in Featurespace’s research, it may be helpful to understand the rationale behind claiming that the results did not support a £2 stake limit. This is, in some sense, purely a semantic argument: while it is undeniable that limiting stakes to £2 would reduced financial loss, a symptom of problem gambling, Featurespace remained committed to the line that they have been asked only to determine if ‘gambling harm’ could be detected in the data. There are a wealth of difficulties with the terminology of ‘harm’ which are addressed in the comments and summary for Report 1. Nonetheless, recognising that the distinction between these two areas of focus and the surrounding semantic acrobatics is useful background to understanding how Featurespace arrived at a brazenly political headline. The overview of Report 1 provided to attendees at the RGT research workshop in December clearly states:

Whilst the aims of this research programme was to attempt to identify harmful patterns of play, most of the evidence reviewed focused on problem gambling. Problem gambling is different from gambling-related harm and we acknowledge this change in focus from the original research objective.

Here is something of a conundrum for those closely investigating the documents. Featurespace’s ‘£2 stake limit’ headline is supported by their claiming that they focused on ‘gambling harm’ and not ‘problem gambling’ is negated by the limitations (above) mentioned in NatCen’s Report 1 overview. Moreover, Report 1 clearly states “There is empirical evidence that suggests that higher levels of staking may be indicative of harm” (page 22). In essence. The rationale for pushing back against the £2 stake limit is founded on an understanding that audiences would not have the time or inclination to fully read the reports as they complement and contradict each other, and therefore the semantics would obfuscate any objections to the claim.

The effectiveness of this headline for Featurespace’s PR cannot be underestimated, as the calculation to back the industry resulted in supportive headlines and a not-so-subtle agreement that

government should turn to Featurespace's analytical methods alone as the solution to problem gambling (J.P. Morgan Casanove, 'Predictive models can identify harmful patterns of gaming'; Business Weekly, 'Cambridge tech could stem the tide of gambling addiction'; as well as eGaming Review and other industry publications warning that operators should not ignore the research and adopt the £2 stake limit, with direct nods to Featurespace's ability to do much better for gaming organisations' bottom line).

Returning to the idea that the research itself was unassailable, the above indicates that there are strong grounds for critiquing the so-called independence of the researchers as well as the £2 stake claim. However, there are other methodological pitfalls which warrant examination.

1. Research is presented in such a way that clearly pushes for the marketing of Featurespace's products rather than objectively describing the results of the analysis. Although there is only one direct mention of Featurespace's trademark 'behavioural analytics' (at the end of report), descriptions of the analytical methodology and interpretation of results are beleaguered by the term 'trade-offs'. This term is used 9 times throughout the report and describes threshold decisions required of those who implement Featurespace's analytical approach.

For example, Figure 4 (page 20) illustrated the way in which thresholds can be massaged to either identify more problem gamblers or avoid inconveniencing more in-control gamblers. These two factors are inversely proportional. In one example, 60% of problem gamblers can be identified and intervened with, if an organisation was willing to also intervene with 30% of players who are not problem gamblers but have been incorrectly identified as such by the model. This is followed by an example of a more accurate model, whereby 60% of problem gamblers could be identified and intervened with, while only 10% of in-control players are incorrectly identified. Finally, if an operator chose another model they could correctly identify 90% of problem gamblers but would need to accept that 30% of in-control players would be incorrectly identified and intervened with.

Operators' obsession with decreasing customer friction by only interacting when it is absolutely necessary undoubtedly had implications for their decided willingness to embrace a software solution that allows them to choose exactly which audience (or which proportion of various audiences) will be receiving intervention. There is an unspoken indication that operators may also determine to not intervene with those they have identified as problem gamblers if their behaviour is at the 'low' end of risk; if customer friction is of a concern for those who are in control of their play, there are even more delicate interactions to be considered when addressing a player who shows signs of addiction but may be adverse to interventions that are perceived as unnecessary or patronising. In essence, it is a key feature of the software that such a granular view would allow operators to determine just how far along the path of addiction individual players are, and set thresholds to intervene when they deem necessary. None of this is relevant in a discussion of independent methodology, but instead more appropriate for marketing materials and sales pitches.

2. PGSI Screening Scores were used as a proxy for 'harmful play.' This screen is widely accepted, but researchers doubted whether it was appropriate for the analysis given its construction (i.e. it is seen as a blunt instrument to measure a complex problem). Featurespace's claim that 'at risk' or 'problematic' behaviours can be identified at all levels of play (that is, evidenced in each metric no matter who the individual is) is not necessarily supported by the PGSI screening score at all scores. Use of the PGSI score produced identification that was 66% more accurate than the baseline ABB Code of Conduct measures, but the models were far more accurate when only the most at-risk players (that

is, those who scored 19 or more on the PGSI score) were examined. More importantly, the use of this score was the most accurate only when lined with the diachronic loyalty card data (550% improvement over the ABB baseline). That combined analysis found that frequency of play was the single most important predictor of problem gambling; however, there are known limitations to using loyalty card data, discussed in both the body of Report 3 and the comments.

Researchers also found that the effectiveness of the PGSI score to predict gambling harm varied by question, with questions 2, 6, 8 and 9 being the most predictive. Featurespace divided these predictive questions into two groups: Question 8 (“How often has your gambling caused you any health problems, including stress and anxiety?”) and Question 9 (“How often has your gambling caused any financial problems for you or your household?”) were deemed ‘more related to gambling related harm’ (Report 3, page 48) than Question 2 (“How often have you needed to gamble with larger amounts of money to get the same feeling of excitement?”) and Question 6 (“How often have people criticized your betting or told you that you had a gambling problem, regardless of whether or not you thought it was true?”). It is not made clear why Featurespace chose to declare that questions 8 and 9 are more related to gambling harm as compared to feelings of excitement vs. staking behaviour (Question 2). When viewed with Adrian Parke et al.’s ‘Stake Size’ study, which details loss of control and feelings of excitement inherent with various staking levels, it appears there is something of an agenda behind recognising that Question 2 is highly predictive but dismissing it as ‘not [as] related to gambling harm’.

3. The use of Proxy Sessions played a key role in researchers’ ability to maximise the data provided. Data from loyalty card players was unavailable for most of the events represented, and therefore models were built to determine what constitutes a ‘session’ of play in order to analyse non-carded gambling. It was not possible to determine a proxy ‘visit,’ and therefore algorithms that predicted when sessions began and ended were used. (Metrics employed to develop this algorithm included machine balance, time between events, and type of event observed.) The limitations to this approach are significant, as the algorithm may identify a proxy session as being shorter than it actually was (reducing the perceived levels of staking, game changing, and so on) or longer than it actually was (leading to a perception of increased levels of spend and duration of play, which will skew the analysis). Play which falls outside of the ‘best fit’ for these models was therefore excluded, providing a very narrow picture for analysis. Although the limitations inherent in this method are implied, there are never fully explained and the researchers state that inaccuracies do not have a significant impact on research findings. Given how proxy session can inaccurately portray staking levels, this appears to be a methodological error with a political edge.

Featurespace’s was, at the time of commissioning, the only known machine learning capacity able to conduct such an analysis. The sheer volume of events alone required computational power that is rarely found outside of universities and research institutions, and indeed Featurespace’s erstwhile connections to Cambridge University are reflected in the robustness of their core analytical engine. An approach that builds individual behavioural profiles and can track and prevent potentially harmful actions is undeniably an important contribution to combatting gambling addiction. It is unfortunate that, despite claims to independence, methodology and results are constructed in such a way as to bolster industry goals and influence legislative decisions rather than advance academic discourse and contribute to external solutions.

Report 4 ‘Patterns of Play; Analysis of data from machines in bookmakers’ forms the last of the ‘core’ research documents for the Responsible Gambling Trust’s B2 Machines Research Programme and is NatCen’s least objective contribution.

Because Report 3 only focused on a subset of the transactional data, ‘Patterns of Play’ was suggested as a way to fill the gap. Unlike the detailed analysis of Report 3’s loyalty card holders’ play and Report 2’s loyalty card customer surveys, ‘Patterns of Play’ was intended to provide a generalised look at play behaviour across the UK. However, at even a general level the results have very little practical use (for example, by using macro divisions of Great Britain rather than a detailed exploration of subsets of cities or towns). This is particularly disappointing given that patterns of play are compared with areas of deprivation and could have provided key information on the percentage of low-income family wealth spent at neighbourhood LBOs.

However, due to the way the betting analysis is segmented, the researchers avoid ever having to admit a grand total amount lost betting on FOBTs for the time covered by the data (6.7 million bets), much less revealing the total loss at a £100 stake. The report also ducks the issue of how net expenditure is distributed by location at a detailed level, providing only a top-layer analysis despite tools and resources readily available for something more granular. This was a deliberate decision to avoid the possibility of providing sensationalist headlines to journalists at the time of publication.

While researchers congratulated themselves on the methodological design, this report indicates how poorly designed the research programme was. That is, Report 4 was designed to provide an industry-positive summary of machine play, but other considerations of methodology were lacking. For example, the analysis Report 4 was conducted in a very slap-dash way (similar to the ‘rapid scoping’ of Report 1). This part of the project was seen as something that needed to be done for appearance’s sake, as the data was there and ready to be used, but that nobody could spend much time on. A detailed analysis would also produce results that would be uncomfortable for Featurespace’s industry customers. There were also political considerations at play, in that the upcoming 2015 election (that is, upcoming at the time of the writing of the report in 2014) could influence policies put forward by the government and therefore any ‘whammies’ that needed to be dropped might be revealed after that time.

Researchers also justified much of their pro-industry stance by touting the low amounts of mean net expenditure their analysis uncovered (e.g. “They may be losing money, but it’s not *that much* money”). However, by acknowledging the influence of B2 games on the amount of money spent (either on B2 games alone or in sessions where B2 and B3 games were combined), researchers might have opened the floor to discussions on whether the product itself is contributing to gambling harm. They were, therefore, careful to avoid blatantly connecting these conclusions.

Although this report was included in the research programme as a document from which policymakers could draw valuable information, the choices made by the authors (reporting the median but not the mean, as elsewhere, or excluding important results of the net expenditure at certain stake levels) render it utterly unsuitable for its intended purpose. Much of the content for Report 4, while generated by Featurespace, was compiled as interpreted by NatCen and is somewhat repetitive of previous reports. Therefore the commentary focuses primarily on the beginning and end of the report, with occasional attention drawn to tables in the body.