Can spending limits in a banking app support safer gambling?

Results from an online lab experiment

December 2023

THE BEHAVIOURAL INSIGHTS TEAM

THE GAMBLING POLICY & RESEARCH UNIT

Contents

Executive summary

Acknowledgement

Section 1: Background

Section 2: Methodology

Section 3: Findings

Appendices

Executive summary.

Experimental design

Spending limits are monetary limits on how much you can spend on gambling through your bank account or card. BIT ran an online lab experiment with 6,049 UK adults who gamble, **testing how four different versions of spending limit tool impact gambling management tool use behaviour.**

Participants were randomised into seeing **five versions of a gambling management tool page in a banking app**. All five versions included a gambling block, and four additionally had different designs of the spending limit:

Control: Gambling block only.
Arm 1: Simple spending limit tool.
Arm 2: Spending limit with a pre-selected £30/month default limit (1% of median UK household income).
Arm 3: Spending limit with £30 default and a message showing what individuals could save based on limit set.
Arm 4: Spending limit with £30 default, and a graph showing previous gambling spend information.

Findings

Offering a spending limit in addition to a block **increased gambling management tool use amongst those who experience no to a low risk** of gambling related harm.

3 in 4 participants would like to see their banks introduce such tools.

The most common reason for using the spending limit tool was to **set a budget.**

There may be a risk of spending limits backfiring for at-risk groups – introducing the tool could lead to those **experiencing moderate or higher levels of gambling harm setting a limit rather than a block.**

Spending limits with a **default limit amount led to individuals setting lower limits on average.**

Comprehension and usability of the spending limit tool was **lower than the gambling block.**

Recommendations

1. Spending limits are a promising *preventative* tool which could help customers manage their gambling spend.

2. We need to consider how to offer spending limits for those who are lower risk, and ensure that other tools are available to those at higher risk. Testing different placements of the tool (e.g., in the budget section) could help to figure this out.

3. Improve and test different designs of the default spending limit interface.

Acknowledgement

The GPRU would like to express thanks to the following individuals and organisations who informed our spending limit tool designs and supported us during this project:

Aisling Ní Chonaire
Esther Hadman
GamCare
The GamCare Lived Experience Community
The Money and Mental Health Policy Institute
Dr Naomi Muggleton
Professor Sharon Collard, University of Bristol
The Co-operative Bank
Nationwide Building Society
Danske Bank
Santander Bank
Lloyds Banking Group
TSB Bank

Section 1: Background

The study's context and aims.

THE BEHAVIOURAL INSIGHTS TEAM

THE GAMBLING POLICY & RESEARCH UNIT

Background - Context

Spending limits could help to reduce gambling related harms, but there is limited evidence on the impact of spending limits in addition to gambling blocks.

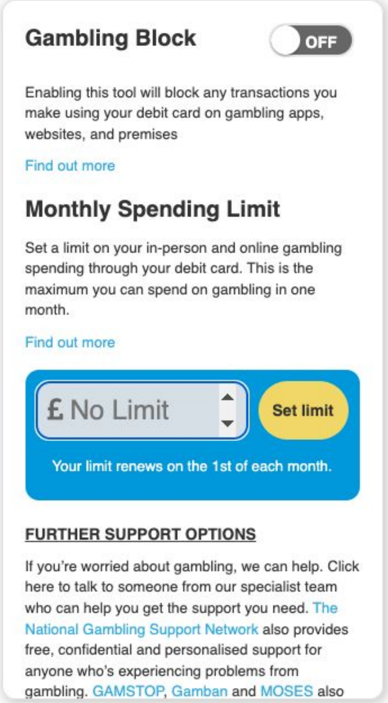


Image of the 'Simple spending limit design'

Gambling spending limits are monetary limits on how much you can spend on gambling through your bank account or card, generally over a month-long period of time.

Banks are well placed to support customers who are at risk of gambling related harm. For example, in a survey with UK residents who gamble frequently online, we found that 6 in 10 deposit funds from a single bank account (n=2,185).[1] Spending limits could thus help customers limit their spending across multiple operator accounts. Whilst most banks offer a gambling block - that is, a tool to stop *all* gambling related transactions - only two UK banks currently offer a spending limit.

However, there is no research on the efficacy of spending limit tools or testing different variations of a spending limit tool. **This study aims to measure the impact of different designs of spending limits on gambling behaviour.**

In an online lab experiment, we tested the impact of the spending limit tool on participants' gambling management tool use, as well as their comprehension of and sentiment towards spending limit tools.

[1] BIT. (2023). Gambling Management Tool Survey (Unpublished)

Our primary research question investigates the potential for spending limit tools to reduce gambling related harm through increased uptake.

RQ1: "Do spending limit tools, when layered on top of gambling blocks, have the potential to decrease gambling harm (compared to gambling blocks alone)?"

We expect that participants who are exposed to a spending limit tool in addition to the gambling block will be more likely to use any gambling management tool, than those who only exposed to the gambling block.

This is based on our previous qualitative research in which spending limits were perceived to be a useful tool and one that people would use if offered by their bank. Our survey also found that 41% said that they were moderately or very likely to use spending limits if offered, compared to the 37% who said they would use a gambling block.[1]

Assumptions: *A spending limit tool has the potential to reduce gambling related harm by reducing spend through increased uptake of gambling management tools, assuming that people who would have otherwise used a block set limits that are low enough to avoid harm.*

[1] Survey data collected by BIT on 19 - 29 January 2022. Publication forthcoming.

Our secondary research question focused on the impact of our spending limit designs on tool use and setting appropriate limits.

RQ2: "Which variation of the gambling spending limit tool is most effective at encouraging people to (a) set a limit and (b) set an appropriate limit?"

We expect that adding behaviourally informed features to a spending limit will (i) further increase the likelihood of participants setting a limit; and (ii) lower the limit set, relative to offering a simple spending limit alone.

Our designs incorporate the following behavioural insights:

- **Defaults:** Our behaviourally informed spending limit tools are all designed to provide a default spending limit to make it easy for participants to set an appropriate limit and anchor them to lower spending limits.
- **Gain framing:** One design also includes a savings message.
- **Salience:** One design additionally incorporates previous gambling spend information.

Our exploratory research questions aimed to understand subgroup differences, customer demand, reasons for using the tool, usability, and comprehension.

RQ3a. "How do results differ across subgroups?"

RQ3b. "Is there customer demand for the spending limit tool?"

RQ3c. "What reasons do people provide for (not) using gambling management tools?"

RQ3d. "Are spending limits tools easy to understand and use?"

We had no clear priors on what we would find in relation to these questions.

However, we expect that there will be differences across subgroups. It could be that people who engage in low to no risk gambling are more likely to use the spending limit tool as a way to prevent an increase in gambling spend or as a precautionary measure. Alternatively, people who gamble may be more likely to use the tool as they enjoy gambling but see the gambling block as too restrictive.

Section 2: Methodology

The study's research design.

THE BEHAVIOURAL INSIGHTS TEAM

THE GAMBLING POLICY & RESEARCH UNIT

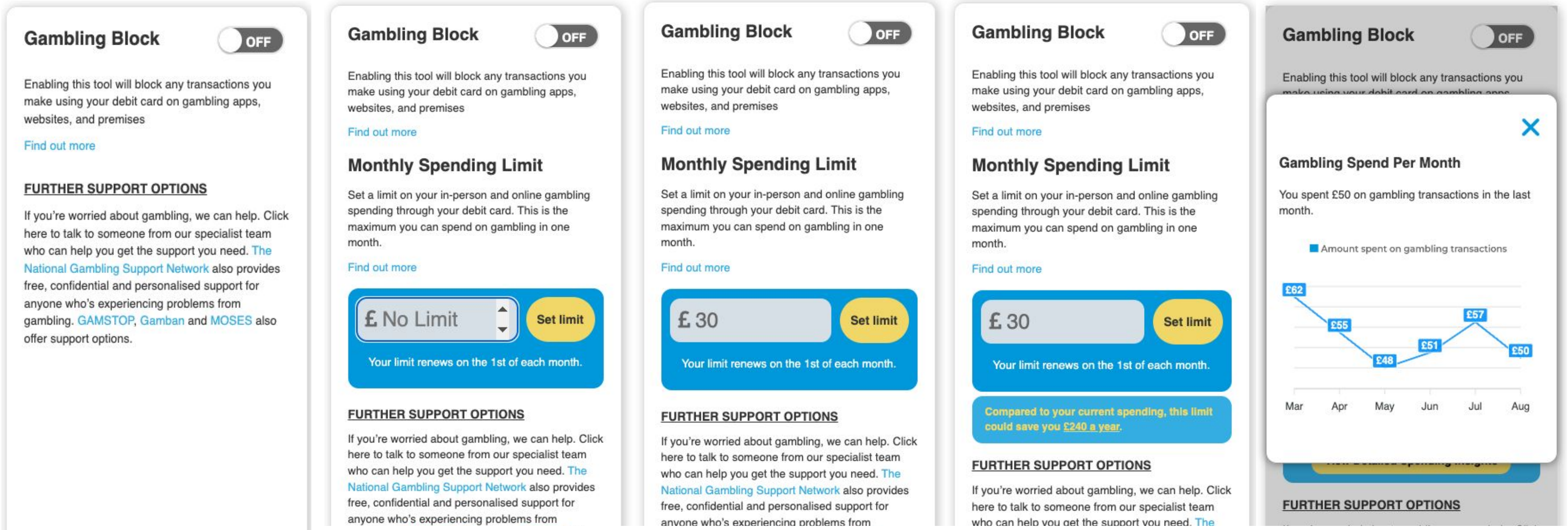
Methodology

Spending Limit Tool Designs

Methodology - Trial arm design

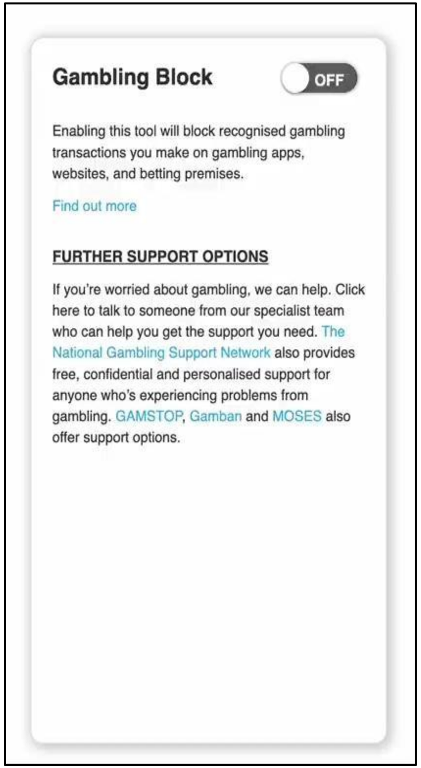
We tested five versions of gambling management tools: the gambling block as the control, and four treatment arms with an additional spending limit tool.

Control: Gambling block	Arm 1: Gambling block + Simple spending limit tool	Arm 2: Gambling Block + spending limit tool with £30 default	Arm 3: Gambling Block + spending limit tool with £30 default + savings message	Arm 4: Gambling Block + spending limit tool with £30 default + spending insights



See Appendix A for more information on the development of the designs. See Appendix B for our Theory of Change which details how we predicted each arm would affect the outcome measures.

Control: Gambling block only.



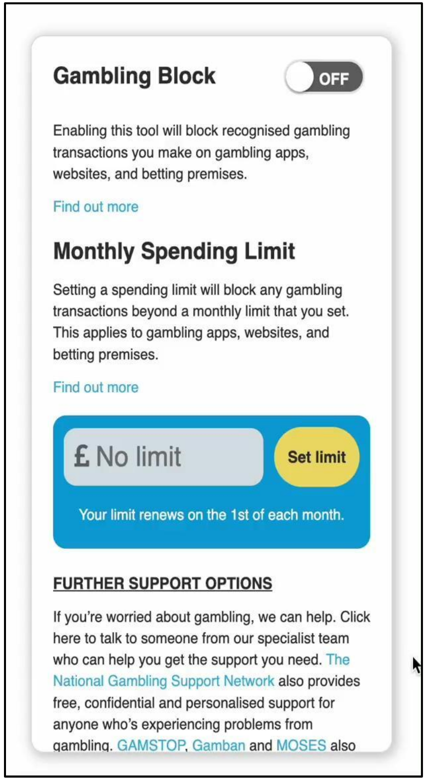
The 'control' version consisted solely of a gambling block, representing what most banks currently offer. The gambling block design was consistent across all treatment arms.

Further support options:
These were added to every treatment arm to signpost where participants could go to receive additional help.

During our co-design workshops, attendees flagged that signposting further support near the tool was missing in our initial design but is crucial for those who feel like they could benefit from further help or who are not aware these other types of support available.

48 hour cooling-off period:
A cooling-off period is the period of time a customer has to wait after disabling the gambling block. We set a 48 hour cooling-off period for the block, as this is what most banks currently offer.

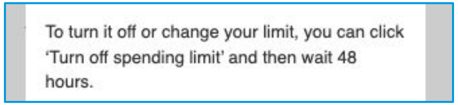
Trial arm 1: Gambling block + Simple Spending Limit.



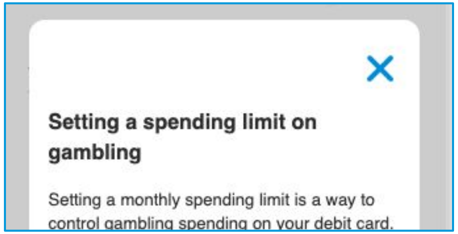
This arm represents a simple spending limit tool, which we could compare the arms with additional behaviourally informed features against. In addition to the gambling block, this arm allows participants to set a limit on how much they can spend on gambling on a monthly basis.

'No limit' default:
Before the spending limit is set, participants are shown that they have 'No Limit' – once a limit is set, this text changes to the amount chosen.

48 hour cooling-off period:
Just as with the gambling block, the spending limit has a 48 hour cooling-off period.

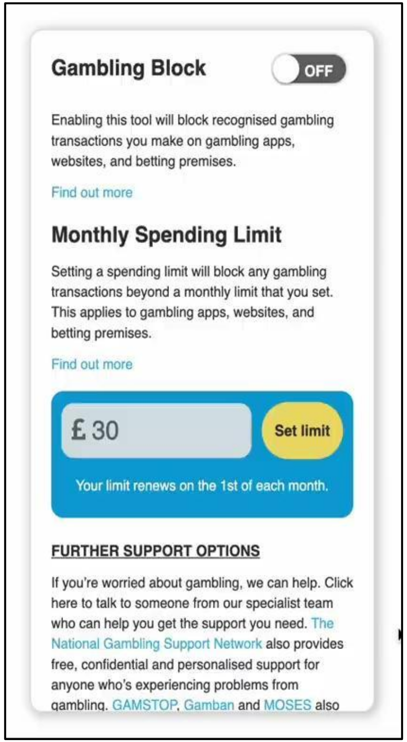


Find out more button:
This and other trial arms have a 'Find out more' button which provides a more detailed description of how the spending limit works.



Methodology - Trial arm design

Trial arm 2: Gambling block + Spending Limit tool with £30 default - This arm incorporates a default spending limit based on safer gambling guidelines.



This arm includes a default spending limit set at £30. This is 1% of median UK household income in 2021.[1] It is based on the Lower Risk Gambling Guidelines (LRGGs) which states that spending 1% or less of household income is associated with low risk of experiencing gambling harm.[2,3]

£30 default:
We set a default amount based on median income because we received feedback from banks that it would be difficult or impossible to tailor a default to an individual's household income.

Defaults: Research shows that defaults (i.e., a pre-selected option) are a very effective way of encouraging people to take a particular course of action, as people often stick to the default.[3]

Anchoring: Research shows that choices tend to be influenced by a suggested reference point.[4] Anchoring participants to the LRGG default may result in participants selecting a spending limit closer to the LRGG.

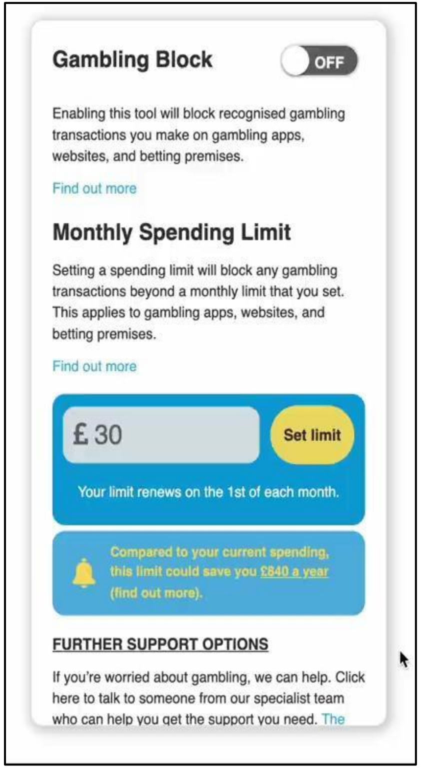
Find out more:
When participants clicked on 'find out more', they were shown additional information regarding the default:

"We've suggested a £30 limit for you based on 1% of median household income in the UK. Research has found that this limit which is 1% of your income, is associated with a lower risk of experiencing gambling harm. Once you reach this limit, we'll block any more gambling transactions made with your card."

[1] The UK median household income for the financial year ending in 2021 was £34,000. 1% of this was divided by 12 and rounded to the nearest 10 to get a monthly limit.
[2] Office for National Statistics. (2022, July 29). Effects of taxes and benefits on UK household income: financial year ending 2021.
[3] Canadian Centre on Substance Use and Addiction. Lower Risk Gambling Guidelines value. This has been replicated in the UK in BIT (2022). https://www.bi.team/publications/lower-risk-gambling-guidelines/.
[4] Jachimowicz, J., Duncan, S., Weber, E., & Johnson, E. (2019). "When and why defaults influence decisions: A meta-analysis of default effects". Behavioural Public Policy, 3(2), 159-186. doi:10.1017/bpp.2018.43
[5] Jung, M. H., Perfecto, H., & Nelson, L. D. (2016). Anchoring in payment: Evaluating a judgmental heuristic in field experimental settings. Journal of Marketing Research, 53(3), 354-368.

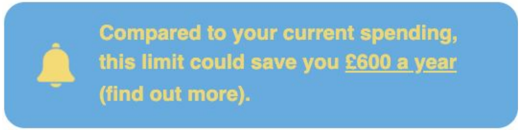
Methodology - Trial arm design

Trial arm 3: Gambling block + Spending Limit tool with £30 default + Savings message - This arm also includes a positive monetary incentive.



In this arm, participants are shown the amount of money they will save in a year, based on the spending limit they set, compared to their current gambling spend.

Savings message:
Initially a yearly savings amount based on the £30 default is shown before the limit is set. When another limit amount is entered, the amount saved adapts based on that value using the following calculation: ***Savings shown = (last month's gambling spend - gambling limit set) x12)***



If a participant enters an amount which is higher than their monthly gambling spending, the message highlights that this is unlikely to reduce their gambling spend.

Positive Incentives & Framing Effects
Framing effect refers to the cognitive bias that our decisions are influenced by the way information is presented.[1] In this arm, this effect is used in two ways:

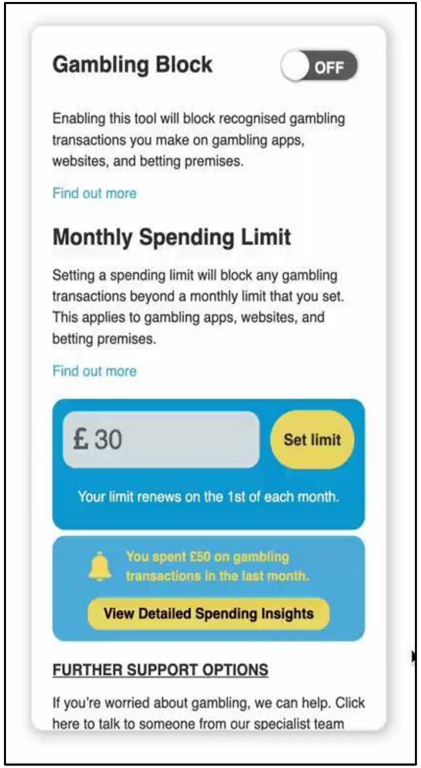
- **Gain framing**: Emphasising positive outcomes by highlighting how much participants could save from setting a limit.
- **Increasing salience and perceived value:** Framing the amount in yearly terms to make the benefits and positive monetary gain from smaller monthly savings more salient and tangible .

[1] Plous, S. (1993). The psychology of judgment and decision making. Mcgraw-Hill Book Company.

Methodology - Trial arm design

Trial arm 4: Gambling block + Spending Limit tool with £30 default + Spending insights

- This arm also includes textual and graphical information on gambling spend.



In addition to the £30 default, this arm also presents information on gambling spend, both under the limit setting button and in a graph if participants click on 'View detailed spending insights'.

Gambling spend message:
Participants are shown how much they spent on gambling in the month.

Gambling spend graph:
Clicking on 'View Detailed Spending Insights' displays a pop up with a graph of their gambling expenditure over the last 6 months. Participants provided an estimate of their previous month's gambling spend so we used that as our 'August' value on the graph, however the other amounts were randomly generated.

Salience

Salience bias is the tendency to focus on items or information that are more noteworthy while ignoring those that do not grab attention.[1] This arm aims to increase the salience and thereby awareness of a participant's gambling spend.

[1] Kahneman D, Slovic P, Tversky A (1982) Judgment Under Uncertainty:Heuristics and Biases(Cambridge University Press, Cambridge,UK)

Methodology

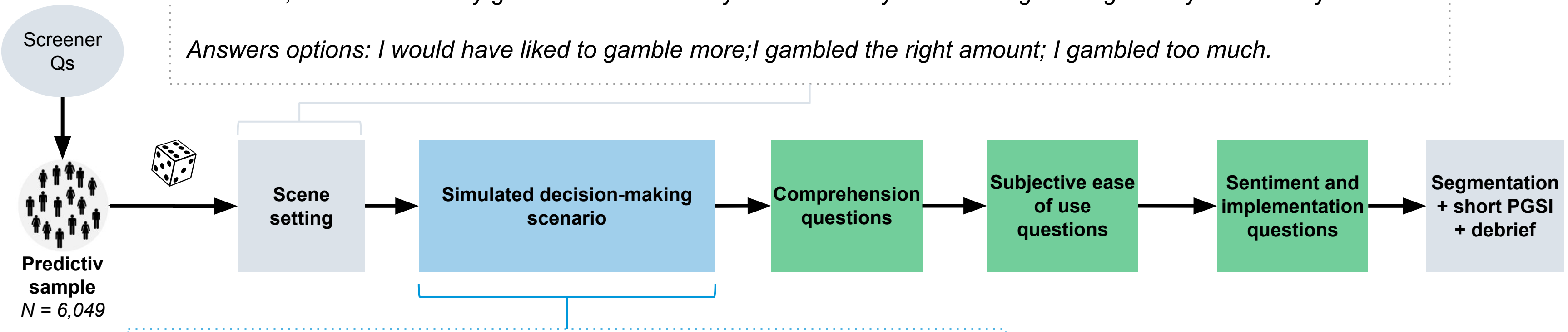
Experimental design

Participants were given a simulated decision-making scenario and asked to use a bank tool page, followed by comprehension, ease of use, and implementation questions.

To encourage 'real world' reactions we 'set the scene' by asking participants:

"Some people are happy with how much they gamble or would like to gamble more. Other people say they gamble too much, and would ideally gamble less. How do you feel about your level of gambling activity in the last year?"

Answers options: I would have liked to gamble more;I gambled the right amount; I gambled too much.



Each participant was given the same scenario to understand how they would interact with the tools in a hypothetical scenario:

"Imagine you have logged onto your mobile banking app that you use to pay for most of your gambling. You are looking through the settings page and find the gambling management page on the following screen.

Please interact with the page as you would if it were actually offered in your bank's app. There will be some follow-up questions."

PGSI = Problem Gambling Severity Index

= Participant randomised to one of five trial arms. We used simple randomisation, with the same likelihood of assignment to each arm.

We recruited a sample of 6,049 participants who regularly gamble in the UK.

BIT recruited 6,049 people who both use mobile banking and had gambled on something other than the National Lottery in the last 12 months to take part in this online experiment. We targeted a nationally representative sample of online banking users who gamble, as defined by our previous research with HSBC.[1]

Education	
Degree	29%

Gender	
Women	39%

Region	
South & East	29%
North	27%
Midlands	17%
Scot/NI/Wales	16%
London	11%

Age	
18-24	17%
25-54	64%
55+	19%

Short PGSI[1]	
Non-risk	51%
Low risk	15%
Med risk	19%
High risk	15%

Ethnicity	
White	86%
Asian	7%
Black	4%
Mixed / other	3%

Median time spent completing the experiment: 6m 36s. We also collected data for all respondents on income, employment and urban location.

[1] BIT (2021). Gambling behaviour: What can bank transaction data tell us? A feasibility study. https://www.bi.team/wp-content/uploads/2021/07/Patterns-of-Play-BIT-HSBC-report-final-June-4th-2021.pdf

N.B. Compared to a sample collected by the Gambling Commission (noting that they use a narrower definition of gambling: individuals who gambled at least once in the past four weeks excluding National Lottery draws), our sample contained more individuals aged 25-54 (30.9% vs 64% in our sample) and more individuals scoring 1 or higher on the short form PSGI (11.2% vs 49% in our sample). The former may be due to one of our eligibility requirements being banking app use. The latter may be due to: a) the GC sample included lotto players, b) it was older on average, and c) the GC survey was administered over the phone, which may have led to participants being less forthcoming in their responses. However, we need to be cautious when extrapolating our results to a wider gambling population.

[1]The short form PGSI is a three question subset of the eight question long-form PGSI.

How to read this report.

Why run an online lab experiment?
We decided to conduct an online lab experiment so we could test variations of the spending limit tool before making suggestions for what to test in a field experiment. An online lab experiment also allows us to test outcomes such as comprehension or sentiment, which might explain impact on behavioural outcomes, such as tool uptake.

What steps did we take to simulate a real world scenario?
- Prioritising objective behavioural outcomes over self-report questions where possible.
- Using a standard UX design to create a ‘gambling management tool’ page. This includes a ‘gambling block’ tool to ensure existing tools offered by banks were presented to participants. This makes the page similar to the page they would typically find in their banking apps.
- Using a neutral scene setting question to encourage participants to consider how they feel about their current gambling behaviour. This is because in the ‘real world’, participants may be likely to search for this page when thinking about their gambling behaviour.

Lab or field?

In lab experiments, participants are presented with information in an artificial and controlled environment and asked to make hypothetical choices. In a field experiment, participants make ‘real’ choices, for example about enabling or disabling a gambling block.

Caveats when interpreting results:
- As this is an online experiment, we should be cautious when extrapolating findings to a real world context. For example, due to the absence of other environmental factors, our results might be stronger than they would be in the real world.
- Our sample size can impact our ability to draw robust inferences. Our sample size was chosen to provide adequate statistical power for our main outcomes of interest between treatment arms. Therefore, we will interpret comparisons for subgroups with caution.

Section 3: Findings

THE
BEHAVIOURAL
INSIGHTS
TEAM

THE
GAMBLING
POLICY &
RESEARCH
UNIT

Findings

Summary of our primary and secondary research questions.

RQ1. Do spending limit tools, when layered on top of gambling blocks, have the potential to decrease gambling harm?

- Offering a spending limit tool alongside a gambling block significantly increased gambling management tool use (6pp - 11pp) compared to when exposed to the gambling block alone.[1]
- However, there is some evidence that spending limits might backfire for ‘at-risk’ groups: introducing the tool could lead to those experiencing higher levels of gambling harm to set a limit rather than a block.

RQ2. Which variation of the gambling spending limit tool is most effective?

- A higher percentage of participants set a limit in the ‘£30 default only’ (25%) and ‘£30 default + savings message’ arm (24%) than the ‘£30 default + spending insights’ arm (20%).
- The £30 default resulted in lower limit setting (~ £91) than the average limit set by those exposed to the simple spending limit (£123). This is driven largely by around 1 in 3 participants sticking to the default limit of £30.
- There were no significant differences in limits set across the three default spending limit arms.

[1] pp stands for ‘percentage points’ which is the difference between two percentages.

Summary of our exploratory research questions.

RQ3.a How do results differ across subgroups?

- When exposed to the spending limit tool, a higher percentage of participants at a lower risk of gambling harm used either tool. However, those at a higher risk of gambling harm were more likely to opt for the limit over the block.
- Among high PGSI participants who set a limit, the £30 default resulted in an average of 52% setting a limit of £30 or less, in comparison to 26% exposed to the simple spending limit tool.
- Financial and gambling literacy did not seem to affect the usage of the gambling management tools.

RQ3.b Is there customer demand for the spending limit tool?

- Approximately 3 in 4 participants would like their banks to provide them with the spending limit tool in addition to a block.

RQ3.c What reasons do people provide for using gambling management tools?

- The most common reason for participants setting a spending limit was to set a budget on their gambling.

RQ3.d Are spending limits tools easy to understand and use?

- Participants stated that both the gambling block and spending limit tool were easy to use and helpful when trying to stick to a budget.
- However, less than 1 in 2 participants answered factual questions about the limit correctly, and 1 in 4 of those who interacted with the tool did not know how to set the limit correctly. This implies that further testing is needed to improve comprehension and usability of the tool.

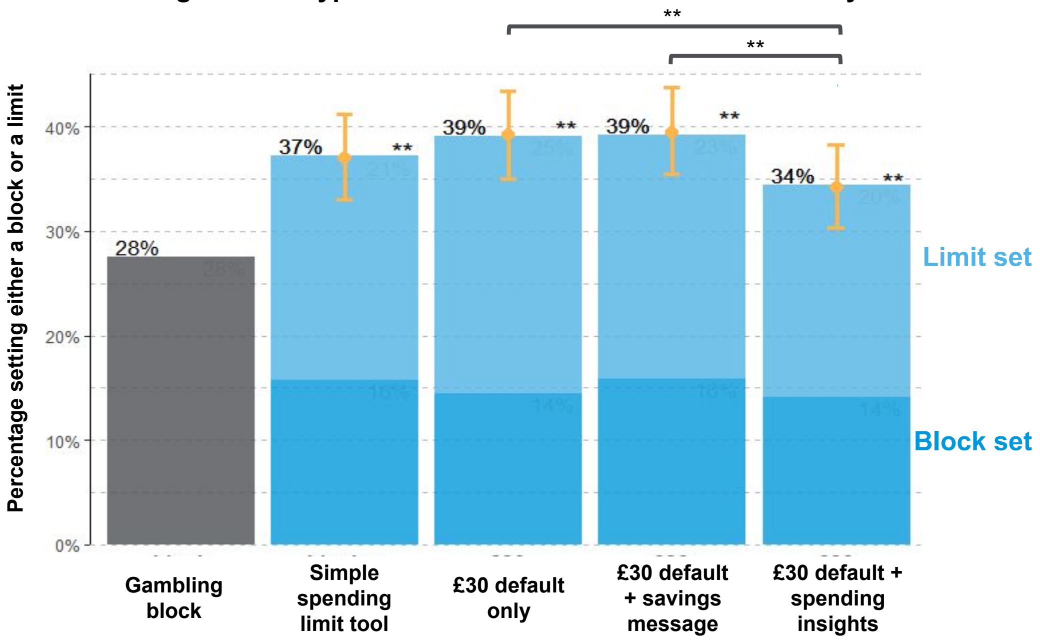
<u>Research Question 1</u>

Do spending limit tools, when layered on top of gambling blocks, have the potential to decrease gambling harm?

RQ1 - Do spending limit tools have the potential to decrease gambling harm?

Offering the spending limit tool increased gambling management tool use. The ‘default only’ and ‘default and savings’ outperformed the spending insights arm.

Percent setting either a hypothetical limit or block the first time they see the tool



Offering a spending limit in addition to the gambling block increases uptake of any tool by 6pp to 11pp (pp stands for ‘percentage points’ which is the difference between two percentages).

The ‘£30 default only’ (39%) and ‘£30 default + savings message’ arm (39%) had a higher uptake than the ‘£30 default + spending insights’ arm (34%).[1]

This indicates that these arms may be more effective at encouraging uptake, and that offering a spending limit means that those who would have otherwise used a block instead set a limit.

[1]£30 default was only larger than the insights graph (p=0.020, adjusted for 10 comparisons), and default with savings larger than spending insights (p=0.014, adjusted for 10 comparisons).

Primary analysis. N = 6,049.

*** p<0.01, * p<0.05, + p<0.10 show covariate-adjusted significance compared to the block only arm after correction for 10 comparisons using the Benjamini-Hochberg Procedure.*

Covariates are PGSI category, age category, gender, region, education dummies, ethnicity, region, and household income above median.

Error bars are 95% confidence intervals for treatment effects versus the block only converted from the log-odds scale. They are not corrected for multiple comparisons.

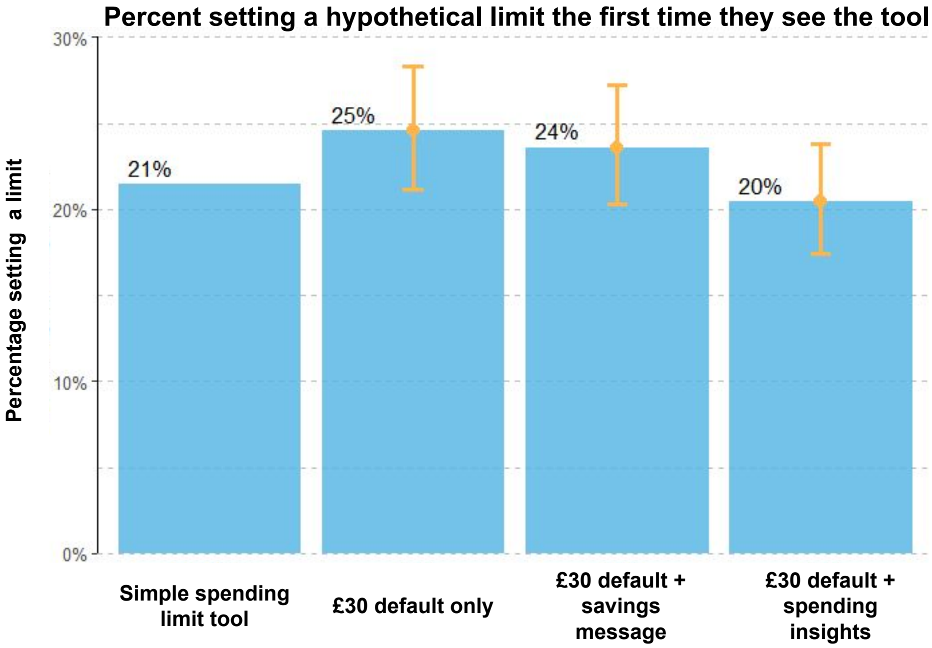
Data collected by BIT on 14th September - 11th October 2023

<u>Research Question 2</u>

Which variation of the gambling spending limit tool is most effective?

RQ2a - Which variation of the gambling spending limit tool is most effective at encouraging people to set a limit?

The ‘default only’ arm led to marginally higher rates of participants setting a limit in comparison to the ‘default and spending insights’ arm.



The ‘£30 default only’ arm (25%) marginally increased limit setting compared to the ‘£30 default + spending insights’ arm (20%) *(p = 0.051)*.[1]

The '£30 default + spending insight' arm was also less easy to understand and less helpful and only 9% of participants in this arm clicked on the graph.

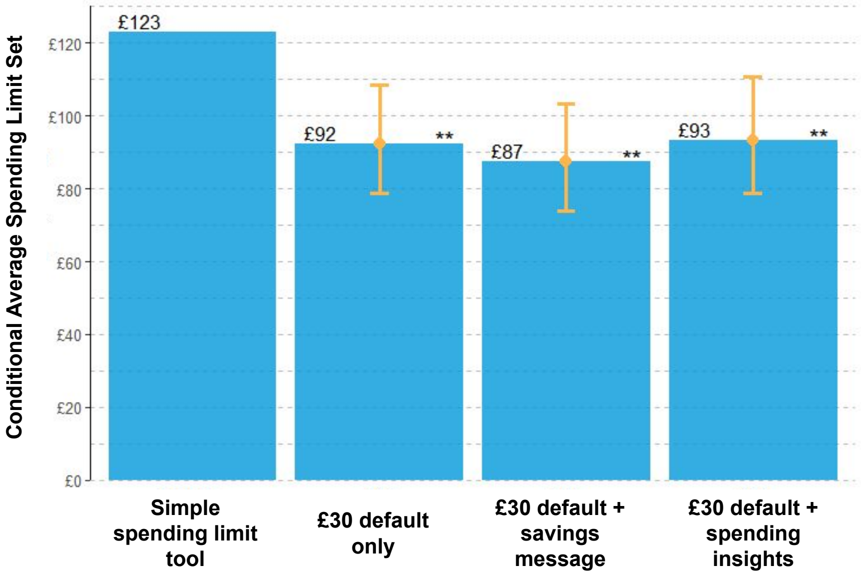
This indicates that further research needs to be conducted on the appropriate design of the spending insights graph, or that spending insights may not provide any additional benefit.

[1]“Default only” had marginally higher rates of limit setting than the “£30 default + spending insights” (p = 0.051, corrected for 12 comparisons)

Secondary analysis. N = 4,862 (excluding block only).
Stars Indicate covariate-adjusted significance compared to the block + limit arm at p<0.05 after correction for 12 comparisons using the Benjamini-Hochberg Procedure.
Covariates are PGSI category, age category, gender, region, education dummies, ethnicity, region, urban area dummy, and household income above median.
Error bars are 95% confidence intervals for treatment effects versus the block only.
Data collected by BIT on 14th September - 11th October 2023

Amongst those who set a spending limit, adding a £30 default significantly reduced the amount that people set their limit at.

Amongst those who set a limit (N=1,086), the average limit set:



Providing a default limit of £30 significantly reduced the amount participants set their limit at.

This is driven by the around 1 in 3 participants that stuck to the default limit of of £30, demonstrating another area in which defaults meaningfully impact behaviour.

There were no significant differences between the limits set in the three arms with a default.

Secondary analysis. N = 1,086.
Those who set a block and those not using any tool are excluded from this analysis.
Stars indicate significance compared to the Block + Limit arm at p<0.05 after correction for 12 comparisons using the Benjamini-Hochberg Procedure.
Covariates include PGSI category, age category, gender, region, education dummies, ethnicity, region, urban area dummy, and household income above median.
Data collected by BIT on 14th September - 11th October 2023.

Research Question 3a

How do results differ across subgroups?

Whilst adding a limit increased tool use among those at lower risk of gambling harm, those at higher risk of gambling harm opted to set a spending limit instead of a block.

	Gambling block	Simple spending limit tool	£30 default	£30 default + savings message	£30 default + spending insights
	No and Low risk (sPGSI 0-1) (N =4,007)				
% of participants who set either a block or limit	22%	36%	39%	39%	34%
% of participants who set a block		14%	12%	13%	13%
	Moderate + High risk (sPGSI 2+) (N = 2,042)				
% of participants who set either a block or limit	37%	40%	40%	39%	35%
% of participants who set a block		20%	19%	21%	17%

Participants who were offered a spending limit were significantly more likely to use a tool than those who only saw a block.

This suggests that the spending limit may unlock a new group of people setting some budgets on their gambling.

Offering a spending limit tool did not significantly increase the use of gambling management tools among those with moderate or high PGSI. This means that when given the choice to set either a spending limit or a gambling block, some higher risk participants are opting for a spending limit when they would have otherwise set a gambling block.

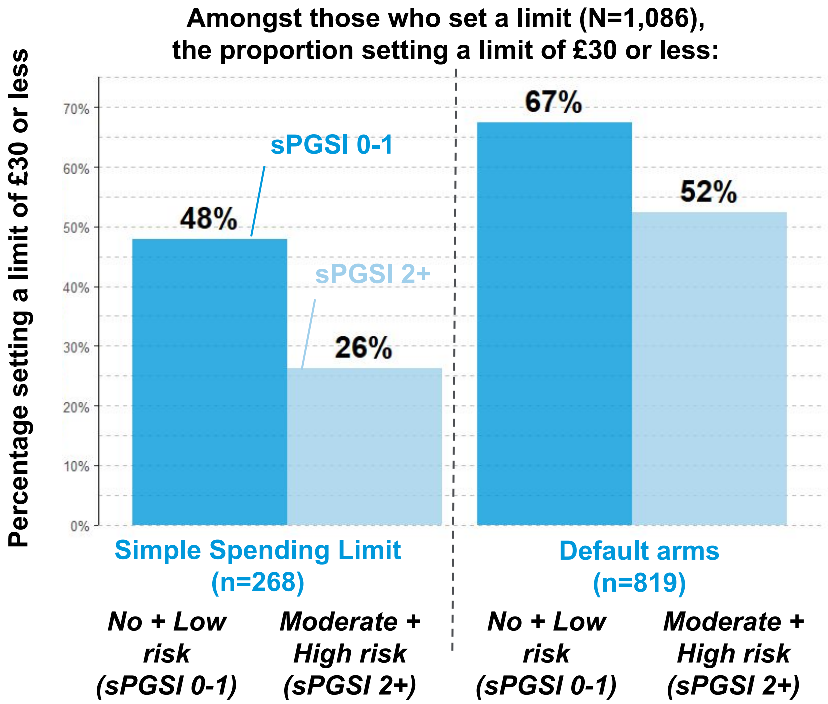
A higher percentage (37%) of moderate to high risk participants enabled the block when only exposed to the block in comparison to no to low risk participants (22%).

We need to consider whether the spending limit tool is the 'right' tool for those at higher risk of gambling harm as more people will choose to keep gambling with a limit when otherwise they might have otherwise stopped gambling.

Exploratory analysis. N=6,049.
Each row is significance tested against the control using Logistic Regression, correcting for four comparisons with the Benjamini Hochberg Method at the 5% significance level.
Cells significantly different that the control are highlighted in green.
Data collected by BIT on 14th September - 11th October 2023

RQ3a - How do results differ across subgroups?

Among those who set a limit, the default doubled the proportion of participants at higher risk of gambling harm setting a limit of £30 or less.



Among high PGSI participants who set a limit, the £30 default resulted in an average of 52% setting a limit of £30 or less, in comparison to 26% when exposed to the simple spending limit tool.

Among low PGSI participants who set a limit, the £30 default resulted in an average of 67% setting a limit of £30 or less, in comparison to 48% when exposed to the simple spending limit tool.

This suggests that defaults may function as an anchor for both those lower and those at higher levels of risk

[1] Overall, 7% of higher risk participants set a limit below £30 in the default arms compared to 5% in the simple spending limit arm.

Descriptive analysis. N = 1,087.

Data collected by BIT on 14th September - 11th October 2023

The impact of the £30 default on the median limit set was stronger for those at higher risk of gambling harm.

Median (mean) limit set by those who do	*No and Low risk (sPGSI 0-1)*	*Moderate + High risk (sPGSI 2+)*
Simple spending limit tool	£40 (£77)	£90 (£222)
£30 default	£30 (£51)	£30 (£287)
£30 default + savings message	£30 (£44)	£48 (£203)
£30 default + spending insights	£30 (£59)	£30 (£158)

The £30 default resulted in a lower median limit set. This change was greater for high PGSI participants, where the median without the £30 default was £90, compared to less than £50 in the £30 default.
The higher average limit amount set by high PGSI participants was sensitive to presence of outliers setting very high limits (nine participants set limits above £1000). We present the full distribution of limits set in these two groups in Appendix D.

Descriptive Analysis. N = 1,087.
See Appendix D for more information on outliers.
Data collected by BIT on 14th September - 11th October 2023

RQ3a - How do results differ across subgroups?

Financial and gambling literacy does not seem to influence gambling management tool use.

	Gambling block	Simple spending limit tool	£30 default	£30 default + savings message	£30 default + spending insights
	Above median financial and gambling literacy (N = 3,197)[1]				
% of participants who set either a block or limit	25%	37%	41%	41%	36%
Mean limit amount set (of those who set one) *(median)*		£146 *(£50)*	£92 *(£30)*	£91 *(£30)*	£99 *(£30)*
	Below median financial and gambling literacy (N= 2,852)[1]				
% of participants who set either a block or limit	30%	38%	37%	37%	33%
Mean limit amount set (of those who set one) *(median)*		£92 *(£50)*	£147 *(£30)*	£79 *(£30)*	£73 *(£30)*

Offering spending limit tools increased uptake among those with lower financial and gambling literacy to almost the same extent as among those with higher literacy.

This is relevant because research indicates that:

- Low financial literacy is associated with future financial harm.[1]
- High levels of gambling fallacies are associated with future gambling harm.[2]

This group therefore might be benefiting most from tools that reduce the risk of harm.

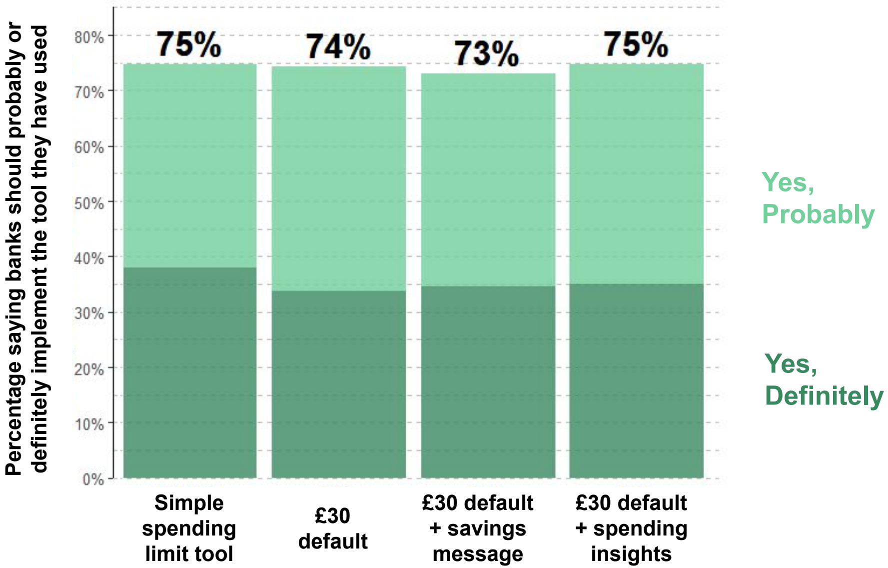
[1] Angrisani, M., Burke, J., Lusardi, A., & Mottola, G. (2023). The evolution of financial literacy over time and its predictive power for financial outcomes: Evidence from longitudinal data. Journal of Pension Economics & Finance, 22(4), 640-657.
Fernandes, D., Lynch Jr, J. G., & Netemeyer, R. G. (2014). Financial literacy, financial education, and downstream financial behaviors. Management science, 60(8), 1861-1883.
[2] Philander, K. S., & Gainsbury, S. M. (2023). An empirical study of the Pathway Model link between cognitive distortions and gambling problems. Journal of Gambling Studies, 39(3), 1189-1205.
Currie, S. R., Hodgins, D. C., Williams, R. J., & Fiest, K. (2021). Predicting future harm from gambling over a five-year period in a general population sample: A survival analysis. BMC psychiatry, 21(1), 1-12..
Participants are defined as having 'high' gambling and financial literacy if they scored above or at the median on a subset of financial literacy (Lusardi and Mitchell, 2014) questions and gambling fallacy questions (Leonard and Williams, 2015). A vulnerable customer is "someone who, due to their personal circumstance, is especially susceptible to detriment, particularly when a firm is not acting with appropriate levels of care." Note that these subscales have not been validated, and the results should therefore be treated as indicative.
Descriptive subgroup analysis, N = 6,049. Data collected by BIT on 14th September - 11th October 2023.

Research Question 3b

Is there customer demand for the spending limit tool?

Overall, around 3 in 4 supported the provision of a spending limit tool.

"Would you like the banks you use to provide a similar spending limit tool in addition to the block?"



A large majority of participants support the introduction of a spending limit tool. *This suggests that adding a spending limit tool and adding a default is just as popular as what banks currently offer which is a gambling block.*

Exploratory analysis. N = 6,049. There were no significant effects in a covariate adjusted logistic regression.
Data collected by BIT on 14th September - 11th October 2023.

Research Question 3c

What reasons do people provide for (not) using gambling management tools?

RQ3c - What reasons do people provide for (not) using gambling management tools?

The most common reason for setting a spending limit was to set a budget on their gambling.

"Why did you decide to set the spending limit over the gambling block?"
Among those who set a spending limit (n= 1,144):[1]

67%	I still want to gamble but want to set a budget
48%	The spending limit better suits my needs
4%	I didn't notice the gambling block
4%	I didn't understand the gambling block
1%	Other reason(s)

The most common reasons participants gave for selecting a spending limit was that it helps with budgeting (67%) and that it meets their needs better than a block (48%).

This result suggests that people may associate the spending limit with budgeting. When testing the placement of the tool within banking apps, we therefore suggest considering placing it in the budgeting section.

A **caveat** to these findings is that participants **might have other reasons for enabling the spending limit** that are not listed in the response options.

[1] Participants could select more than one reason;
Descriptive analysis. N = 6,049.
Data collected by BIT on 14th September - 11th October 2023

The most common reason for not using either tool was that participants perceived that they did not need to restrict their gambling.

“What are your main reasons for <u>not</u> setting a block or limit?”
Among those who set neither a block nor limit (n=3,042)[1]:

59%	I don’t need to restrict my gambling		
14%	This isn’t “real” so my actions didn’t matter		
11%	I don’t think either tool would work for me		
7%	I don’t understand the tools	7%	I didn’t notice the block
6%	I didn’t notice the limit	6%	I don’t understand the block
5%	I don’t understand the limit	5%	Other reason(s)

44% of participants who engage in high-risk gambling said they did not need to restrict their gambling, vs 67% of those who engage in low-risk gambling.

A **caveat** to these findings is that participants **might have other reasons for not using the spending limit** that are not listed here and they chose to enter these in the ‘other reasons’ option.

While the most common reason for not using a tool was that individuals did not perceive that they needed to restrict their gambling (59%), some participants also indicated that they did **not understand the tool, notice the tool, or found it difficult to use**. *This suggests that further research, such as a field trial, could look into how to improve users’ understanding and recognition of these tools to drive uptake.*

The second most common reason cited for not using a tool was that individuals were in a simulated environment rather than a ‘real’ one (14%). *This suggests that the usage of gambling management tools may be higher if individuals are exposed to the tool in a field trial / within their actual banking app.*

[1] Participants could select more than one reason;
Descriptive analysis. N = 6,049.
Data collected by BIT on 14th September - 11th October 2023

RQ3c - What reasons do people provide for (not) using gambling management tools?

The most common reason to enable a gambling block was that participants saw it better suiting their needs.

“Why did you choose to set a block and not a limit?”
Among those who set a block (n=733)[1]

50%	The gambling block better suits my needs
28%	I want to cut back my gambling spend
20%	The spending limit wouldn’t help me stick to a budget
17%	I didn’t notice the spending limit
6%	I didn’t understand the spending limit
2%	Other reason(s)

Of those who chose to set a gambling block over a limit, the most common reason cited was that the block suited their needs better (50%).

However, some participants did not notice or understand the spending limit tool.

This suggests that there is a need to improve the usability of the spending limit tool to increase uptake.

Out of those who selected ‘the gambling block better suits my needs’ and ‘I want to cut back my gambling spend’ there were no significant differences between those at a lower risk of gambling harm compared to those at a higher risk.

[1] Participants could select more than one reason;
Descriptive analysis. N = 6,049.
Data collected by BIT on 14th September - 11th October 2023

Research Question 3d

Are spending limits tools easy to understand and use?

RQ3d - Are spending limits tools easy to understand and use?

The gambling management tools were seen as easy to use and helpful when trying to stick to a budget.

"To what extent do you feel the gambling management tool you just used is…"

% saying the tool was moderately or very…	Gambling block	Simple spending limit tool	£30 default only	£30 default + savings message	£30 default + spending insights
N	1,187	1,248	1,146	1,225	1,243
…easy to use?	87%	90%*	85%	84%+	83%*
…helpful when trying to stick to a gambling budget?	82%	88%*	86%*	84%	84%

Overall, more than 4 in 5 found the tools easy to use and helpful.

Compared to those who saw the gambling block only, **more participants** said the tool was moderately or very easy to use when using **the simple spending limit, while fewer** said the same when they **saw the '30 default + spending insights'** arm.

Adding a spending limit **increased the proportion of people thinking the tool would be helpful** when trying to stick to a budget. The largest increases were in the '£30 default only' and 'simple spending limit' arms.

Importantly, adding in a default does not have a backfire effect compared to the block only – i.e., participants did not perceive it as being more challenging to use.

Exploratory analysis. N = 6,049.
Comparisons are made between each treatment arm and the gambling block arm.
** Indicates p<0.05 in a covariate adjusted logistic regression when corrected for four multiple comparisons (within each row). + indicates p < 0.10.*
Covariates include short-form PGSI above 2+, age category, gender, region, education dummies, ethnicity, region, urban area dummy, and household income above median.
Data collected by BIT on 14th September - 11th October 2023

While 4 in 5 participants understood what the gambling block does, less than 1 in 2 understood how a spending limit works.

When asked what the gambling management tools do…

% who correctly said…	Gambling block	Simple spending limit tool	£30 default	£30 default + savings message	£30 default + spending insights
N	**1,187**	**1,248**	**1,146**	**1,225**	**1,243**
The **gambling block** prevents the user from making any gambling transactions	79%	82%	79%	81%	80%
The **spending limit** blocks any gambling transactions beyond a monthly maximum	-	45%	46%	48%	45%

Overall, comprehension of what the gambling block does was high among participants (approx. 80%).

In comparison, comprehension of what the spending limit is lower – less than half of the participants knew what the limit does correctly.

This may be because spending limit tools are inherently more complex, or because the design we used was not optimal.

Alternatively, it may be due to the fact that few banks currently offer the spending limit tool compared to the block.

This suggests that further testing needs to take place to improve the design of the tool.

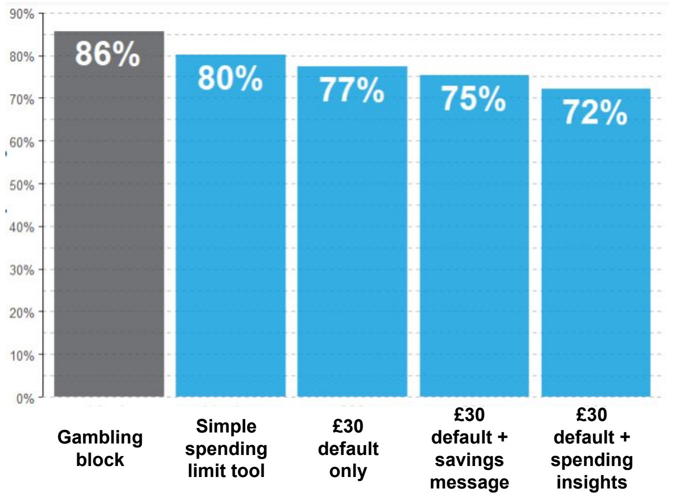
Exploratory analysis. N = 6,049.
Comparisons are made between each treatment arm and the gambling block arm.
** Indicates p<0.05 in a covariate adjusted logistic regression when corrected for four multiple comparisons (within each row). + indicates p < 0.10.*
Covariates include short-form PGSI above 2+, age category, gender, region, education dummies, ethnicity, region, urban area dummy, and household income above median.
Data collected by BIT on 14th September - 11th October 2023

When asked to use the management tools in a certain way, of those who interacted with the tool, 4 in 5 set the block correctly, and 3 in 4 set the limit correctly

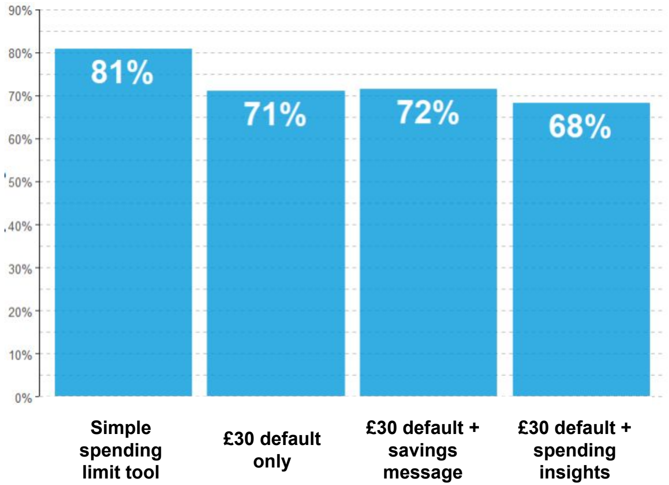
When instructed to turn the gambling block on, 78% of participants who interacted with the tool correctly set the block. Accuracy was lower in the arms with a limit.

When instructed to set a spending limit of £80, 73% of participants who interacted with the tool correctly set the limit. The 'simple spending limit' arm outperformed the other default arms.

% of those interacting who set the block correctly **(n=5,079)**



% of those interacting who set the limit correctly **(n=3,669)**

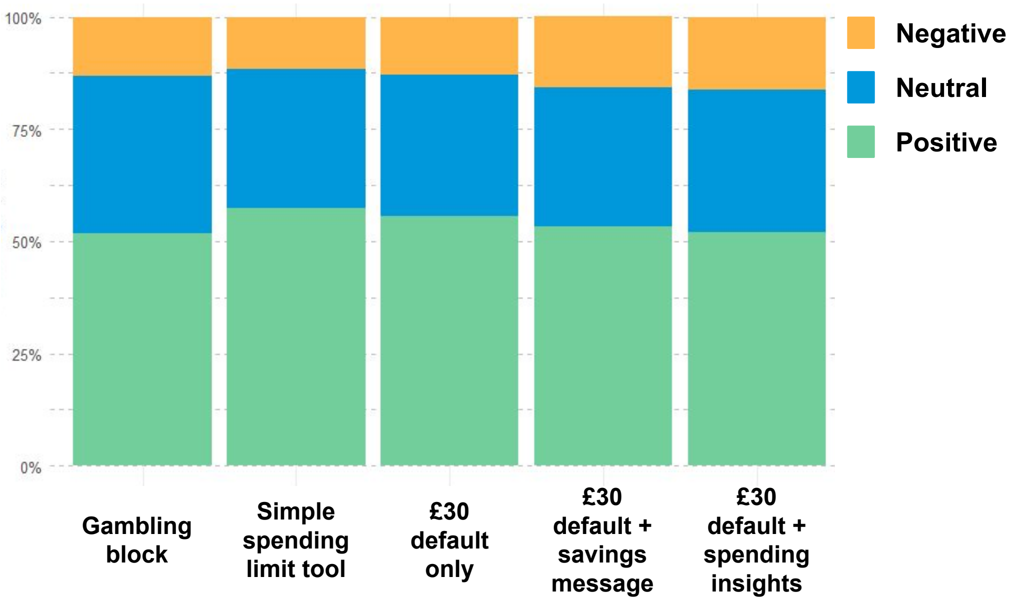


Here, we exclude participants who did not interact with the tool based on the assumption that they wanted to progress with the survey rather than engage with the task. When including all participants, we found that the percentage of participants using the tool correctly was much lower. This analysis is reported in Appendix C.

Descriptive analysis. Out of those who did not interact with the tool, we found that many did not interact with any tool feature, suggesting lower data quality. The amount of time to get a correct answer was the same across treatment arms. We investigated a potential usability issue or glitch with the simulation as the cause for the low percentage of participants using the tool correctly. We also conducted various data quality checks and our own usability tests.
Data collected by BIT on 14th September - 11th October 2023.

Across arms, the majority of the feedback on the tool was positive.

% of participant responses coded as positive, neutral or negative (n=2,164)



For each treatment arm we conducted AI-assisted sentiment analysis on responses to the following optional question: “Do you have any other feedback on the design of the gambling management tool?”.[1]

Sentiment analysis consists of analysing text to determine whether the emotional tone of the message is positive, negative, or neutral.

Across all arms, the proportion of feedback that was positive, negative, or neutral was similar, with more than 50% of responses being rated as positive.

[1]OpenAI. (2023). *ChatGPT* [Large language model]. https://chat.openai.com/chat.

Sentiment analysis. Out of those who responded to the following optional survey question “Do you have any other feedback on the design of the gambling management tool?” we conducted exploratory sentiment analysis assisted by a large language model (LLM). See Appendix E for details on our methodology.

Participant feedback suggests that there is scope to iteratively improve the design and functionality of the tool.

For each arm we conducted AI-assisted categorisation of responses (n= 1,539) to the following optional question: “Do you have any other feedback on the design of the gambling management tool?”.[1] The responses could be summarised under four categories:

The overall tone of the feedback was **mostly positive** and supportive of the proposed tool, with many users finding the spending limit tool straightforward and easy to use.	*"This would be really helpful for someone who struggles to budget and control their gambling." [£30 default only]* *"I think it's a good idea and its simplicity should be praised." [£30 default & spending insights]*
Some expressed doubts about the **effectiveness of the tool**, particularly for those at higher risk of gambling related harm.	*"Useful but would people with serious issues use the tool properly?" [£30 default only]* *"It's a good idea, but would an addict use it" [Simple spending limit]*
Some participants mentioned that certain features were confusing and that they had **technical difficulties**.	*"Couldn't work out how to set the limit at £80, it defaulted to £30." [£30 default & spending insights]* [2] *"Just confusing to use. It explains okay but to physically use is poor." [£30 default only]*
Participants provided **suggestions for improvement** related to the design, functionality, and features, amongst others.	*(see page 47 for examples)*

[1]OpenAI. (2023). *ChatGPT* [Large language model]. https://chat.openai.com/chat. We used an LLM to initially categorise the responses before subsequently manually drawing out key categories.

[2] See page 44 for additional statistics on usability testing.

Out of those who responded to the following optional survey question: “Do you have any other feedback on the design of the gambling management tool?” we categorised and summarised responses assisted by a large language model (LLM). See Appendix E for details on our methodology.

Across all treatment arms, participants provided suggestions for improvement.

For each arm, we conducted further AI-assisted categorisation of responses (n=1,539) to identify suggestions for and categories of improvements. We then manually identified similarities and differences in categories across arms, leading to a consolidation of four areas of improvement.

User interface and design	*“Print a little bit clearer at the bottom of the page as likely to be ignored as text too small" [£30 default only]* *“There is a lot of writing on this page which could be a tad overwhelming. If this was simplified down to the just-needed text, it would avoid looking so busy” [£30 default & spending insights]*
Ease of use and clarity	*“The instructions need to be slightly clearer, I thought pressing 'set limit' would then give the option to choose the limit amount.” [£30 default only]* *"[The] choice between block or set limit should be more obvious initially. there’s also too much text to read.” [Simple spending limit]*
Cooling-off period	*“I think there should be an option to block indefinitely, and maybe you have to contact the bank to unblock.” [Simple spending limit]* *“Make it more than 48 hours because i can just turn it off and be fine in 2 days but for a week it’s more serious” [£30 default & savings message]*
Additional features	*“Being able to change the renewal date would be handy, so i can align to my pay date” [£30 default & savings message]* *"Perhaps some advice on limits and 'what is a healthy limit' for people to get an understanding of the concept.” [£30 default]*

[1]OpenAI. (2023). *ChatGPT* [Large language model]. https://chat.openai.com/chat. We used an LLM to initially categorise the responses before subsequently manually drawing out key areas for improvement.
Out of those who responded to the following optional survey question: “Do you have any other feedback on the design of the gambling management tool?” we categorised and summarised responses assisted by a large language model (LLM). See Appendix E for details on our methodology.

Appendix A - Background

THE BEHAVIOURAL INSIGHTS TEAM

THE GAMBLING POLICY & RESEARCH UNIT

Development of spending limit tools to test.

We went through a multi-stage research process to (1) identify what type of gambling management tool that can be offered by financial services firms to test; (2) refine the chosen tool.

Objective	Activities
Understand existing tools and associated research	• Evidence review on the availability and effectiveness of gambling support tools via financial services firms. • Interviews with a range of key stakeholders, including banks, research organisations, and academia. • A review of the gambling support tools available via major retail banks. • Survey and interviews with people who gamble to understand the awareness and appetite for tools, as well as trust in banks offering support.
Decision to test a spending limit.	
Refine spending limits tools	• We ran two co-design workshops to acquire feedback on three designs of the spending limit tool from banks and those with lived experience of gambling related harm. • We conducted a trial arm ranking exercise which included 13 potential trial arms. Six BIT colleagues and one external expert took part in this exercise. • We requested feedback from banks, lived experience, and academics on three of our treatment arms. We consolidated feedback and used this to inform final tweaks to our designs.

Appendix B - Theory of Change

Our Theory of Change outlines how we predicted each 1) arm would affect the 3) outcome measures, through the 2) mechanisms.

(1) Arm

These are the spending limit tools we designed to measure the impact of gambling tool use.

(2) Mechanisms

These are the causal chains through which we think the link between spending limit tool features and our outcome measures work. They include several links which run from engaging with the tool to the outcome measures we are interested in.

(3) Behavioural outcome measures

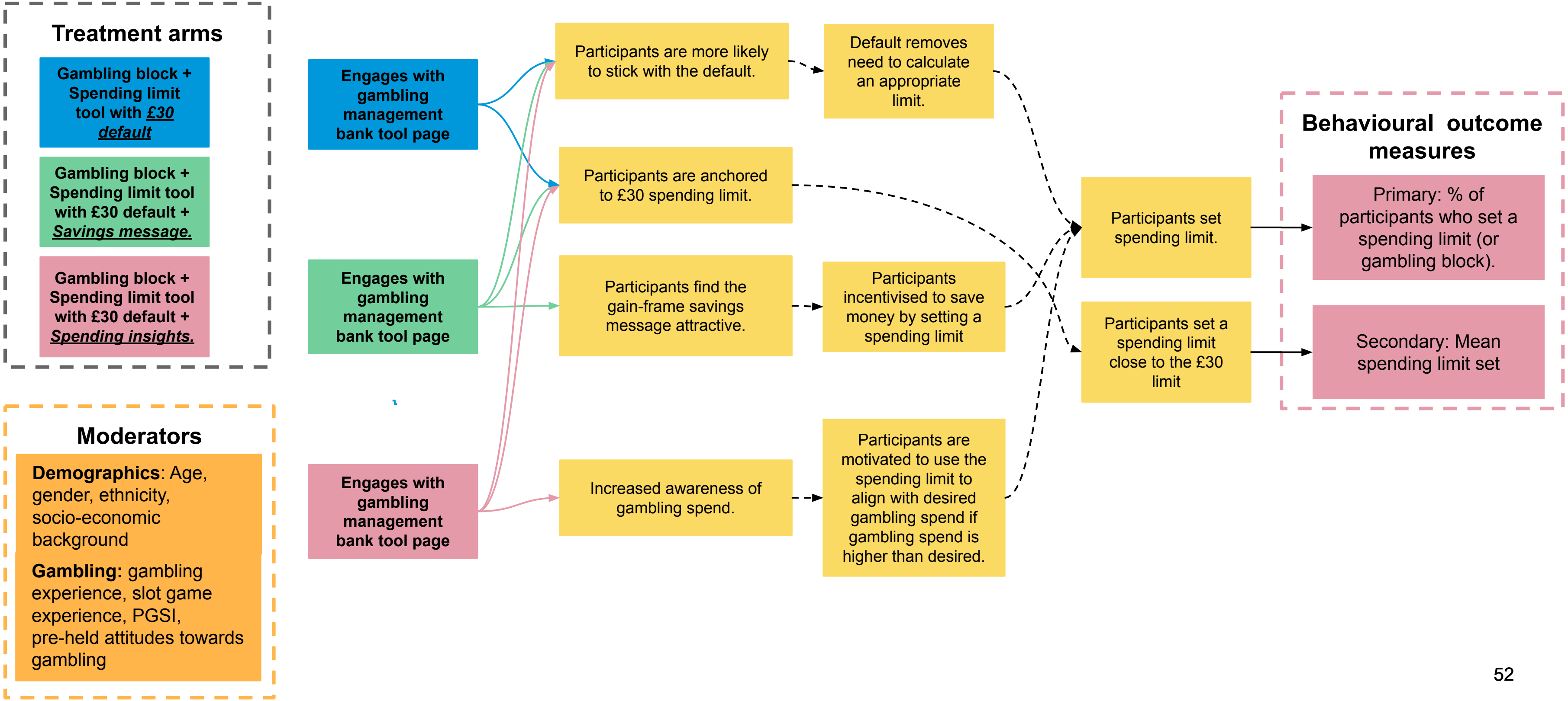
This set of outcome measures examines outcomes related to participants' behaviour in the simulated decision making scenario.

Moderators

These are factors that could potentially have an influence on the impact of spending limit tools on our outcome measures.

Appendix B: Theory of Change

We identified that spending limits incorporating behavioural insights to be the key factor in affecting gambling behaviour.



Appendix C - Supplementary results

THE BEHAVIOURAL INSIGHTS TEAM

THE GAMBLING POLICY & RESEARCH UNIT

Balance checks: there was good balance across treatment arms for our covariates.

Data collected by BIT on 18 May - 12 June 2023.	Gambling Block (N=1,187)	Simple spending limit (N=1,248)	£30 Default (N=1,146)	£30 Default + Savings Message (N=1,225)	£30 Default + Spending Insights (N=1,243)
Mean Age (years) *(mean, (sd))*	40 *(15)*	39 *(14)*	40 *(15)*	41 *(15)*	40 *(14)*
Gender (female) *(count, (%))*	487 (41%)	477 (38%)	440 (38%)	466 (38%)	484 (39%)
Mean Short PGSI Score (0-9 scale) *(mean, (sd))*	1.5 *(2.1)*	1.4 *(2.1)*	1.4 *(2.0)*	1.5 *(2.1)*	1.5 (2.1)
Higher-risk gambler (sPGSI 2+) *(count, (%))*	404 (34%)	411 (33%)	368 (32%)	425 (35%)	434 (35%)
Mean Number of gambling types (last 12 months) *(mean, (sd))*	5.7 *(2.9)*	5.7 *(2.9)*	5.7 *(2.9)*	5.7 *(2.9)*	5.8 *(3.0)*
Employed (part-time or full-time) *(count, (%))*	893 (75%)	942 (75%)	883 (77%)	914 (75%)	933 (75%)
Education (has degree or higher) *(count, (%))*	347 (29%)	346 (28%)	330 (29%)	374 (31%)	361 (29%)
Income (above £40k) *(count, (%))*	533 (45%)	557 (45%)	533 (47%)	596 (49%)	579 (47%)
Ethnicity (non-white) *(count, (%))*	179 (15%)	177 (14%)	163 (14%)	179 (15%)	175 (14%)
Region (not in England) *(count, (%))*	203 (17%)	195 (16%)	183 (16%)	177 (14%)	191 (15%)
Felt gambled too much (last 12 months) *(count, (%))*	113 (10%)	128 (10%)	94 (8%)	109 (9%)	123 (10%)
Used gambling tools before (asked after experiment) *(count, (%))*	326 (27%)	345 (28%)	298 (26%)	325 (27%)	369 (30%)

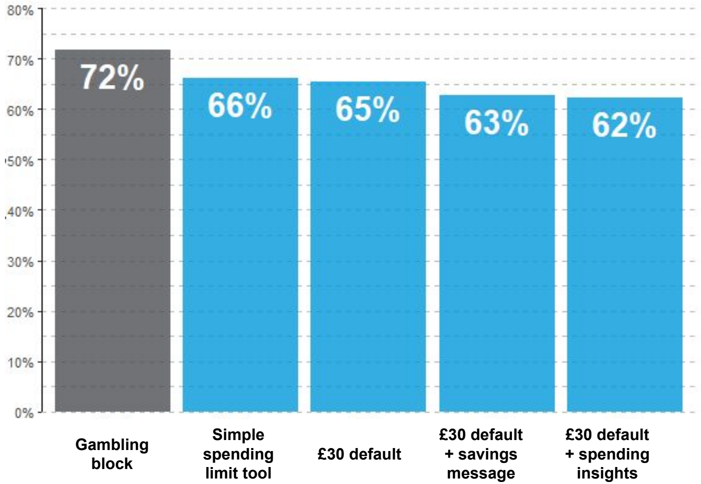
Italicised values in brackets are standard deviations.

Including all participants, during user testing 2 in 3 set the block correctly when and just over half set the limit correctly.

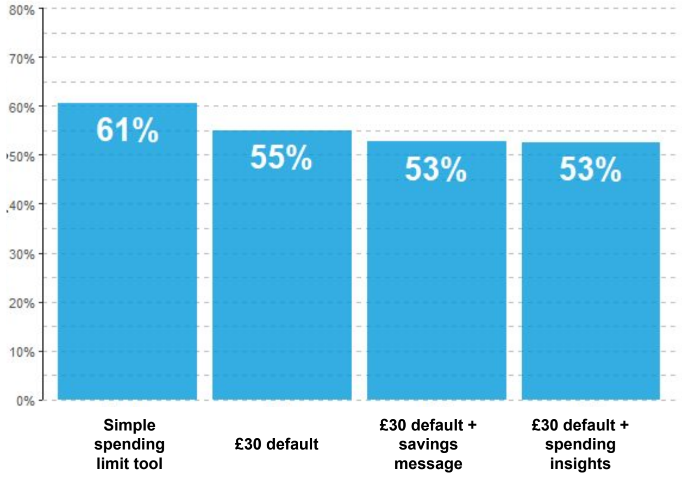
When instructed to turn on the gambling block as a user test, ~66% of participants correctly set the block. 970 participants did not interact with any of the tool's features.

When instructed to set a spending limit of £80, 55% of participants correctly set the limit. 2,380 participants did not interact with any of the tool's features. Lack of interaction could be higher as it was the second usability test.

% Setting the block correctly (n=6,049)



% Setting the limit correctly (n=6,049)



Descriptive analysis. N=6,049.
The amount of time to get a correct answer was the same across treatment arms.
Data collected by BIT on 18 May - 12 June 2023

Appendix D - Spending Limit Set: Outliers

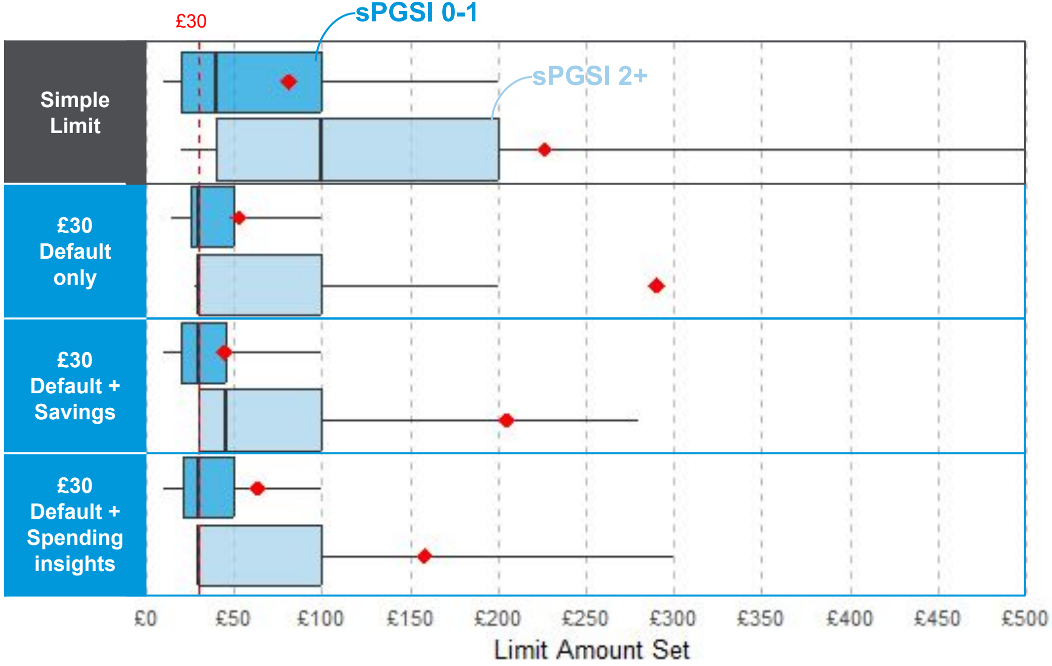
THE BEHAVIOURAL INSIGHTS TEAM

THE GAMBLING POLICY & RESEARCH UNIT

Appendix D - Spending Limit Set: Outliers

The higher average spending limit amount set by high PGSI participants is mainly due to presence of outliers.

Distribution of limits set split by sPGSI category



For higher PGSI participants, the means are consistently above the 75th percentile of the data. This is partly driven by outliers, i.e., a small number of participants setting very high limits (e.g. nine participants set limits above £1000) that have an outside impact on the mean.

We decided against removing these outliers for quantitative analysis as we believe that they form an important part of the story - indicating that a small number of individuals might be setting unreasonably high limits. Winsorized results are presented on the next page.

Reading the plot:

- **Red** dots are means.
- **Whiskers** (e.g. the black lines) reach from 10th percentile to 90th percentile.
- The **blue** boxes have three lines. From left to right: 25th percentile, 50th percentile (median), and 75th percentile.

N = 1,087. Data collected by BIT on 14th September - 11th October 2023. Differences may also be due to different sample characteristics.

A small number of very high spending limits set were impacting the results of high PGSI participants.

	Gambling block	Simple spending limit tool	£30 default only	£30 default + savings message	£30 default + spending insights
	No and Low risk (sPGSI 0-1) who set a limit (n=770)				
Average (median) limit set by those who do	-	£77 (£40)	£51 (£30)	£44 (£30)	£59 (£30)
Winsorized average limit set by those who do	-	£77	£51	£44	£59
	Moderate + High risk (sPGSI 2+) who set a limit (n=317)				
Average limit set by those who do	-	£222 (£90)	£287 (£30)	£203 (£48)	£158 (£30)
Winsorized average limit set by those who do	-	£174	£120	£123	£120

Winsorizing maps spending limits set above certain percentile to this percentile. The purpose is to limit extreme values in the data to reduce the effect of outliers disproportionately impacting the results. In this case, any values above £1,000 (the 99th percentile) were assigned a value of £1,000. For example, if someone set a limit of £5,000, we would have assigned a value of £1,000. There were 9 data points over £1,000.

The sensitivity of the means in the sPGSI 2+ group to winsorizing 1% of the data illustrates how influential these higher limit setters are.

We have decided not to take this approach in the main body of the analysis for two reasons. First, the approach to winsorizing is arbitrary (i.e. picking a value of £1,000 is subjective, we could have just as easily decided on £500 or £2,000). Second, these are important data points when understanding the potential for gambling related harm.

N = 1,087. Data collected by BIT on 14th September - 11th October 2023

Appendix E - Free text categorisation

THE BEHAVIOURAL INSIGHTS TEAM

THE GAMBLING POLICY & RESEARCH UNIT

We used a large language model to conduct sentiment analysis on free-text feedback on the design of our tools.[1]

Goal	To understand the emotional tone of free-text survey responses by categorising responses into three categories: positive, neutral and negative sentiment.
Sample	After removing personal data (see page 63), we also removed responses amounting to a 'no' or 'nothing to add'. The final sample included 2,164 free-text responses.
Method	**Sentiment analysis:** 1. We used a script in Python which takes a csv of survey data (one row per participant) and asks a large language model (LLM) to classify each response (one piece of feedback per participant), without any training. 2. We then asked the LLM to calculate the percentage of feedback for each type of sentiment per arm, to generate examples for each category, and plot the results for each treatment arm. 3. We used the percentages to create a stacked bar chart to visually show the proportion of each type of sentiment by arm (see page 45).

[1]Note that we used the following large language model (LLM) for all analysis: OpenAI. (2023). *ChatGPT* (Dec 7 version) [Large language model]. https://chat.openai.com/chat

We used a large language model to categorise free-text feedback on the design of our tools in order to identify key topics.[1]

Goal	To summarise and categorise free-text responses into distinct topic areas for each arm so we could identify differences, similarities, and patterns across arms.
Sample	After removing personal data (see page 63), we also removed responses that amounted to a 'no' or 'nothing to add' response. We also removed responses with three words or less, leaving us with a final sample of 1,539. We split responses into treatment arms (~300 per arm) and then split the treatment arm data into two random datasets (~150 per dataset) because the large language model (LLM) could not analyse all 300 response in one go.
Method	**AI-assisted categorisation of responses (for each arm)**: 1. Ran prompt A: we used a script in R to ask an LLM to generate 4-6 categories for the first ~150 pieces of feedback. As we had already manually identified that the feedback included suggestions for improvements (see page 64) and we wanted to understand how to improve our tool design, we explicitly asked the LLM to include suggestions for improvement as one of the categories. 2. Ran prompt B: we used a script in R to ask an LLM to sort all the remaining ~150 responses into the 4-6 categories generated by prompt A. Across all arms this totalled 28 categories. 3. We then manually looked at the output to identify similarities and differences in categories within and across treatment arms. Since the categories within and across treatment arms were similar, we organised them into four overarching categories across all treatment arms (see page 46).

[1]Note that we used the following large language model (LLM) for all analysis: OpenAI. (2023). *ChatGPT* (Dec 7 version) [Large language model]. https://chat.openai.com/chat

We used a large language model to further identify and categorise suggestions for improvement.[1]

Goal	To understand how to improve the design of the spending limit tool by summarising and grouping free-text responses into specific areas of improvement for each arm, as well as identifying differences, similarities, and patterns across arms.
Sample	We used the same sample of 1,539 responses and method that we used on page 61 in which we split responses into treatment arms (~300 per arm) and then split the treatment arm data into two random datasets (~150 per dataset) because the large language model (LLM) could not analyse all 300 response in one go.
Method	**Identify and categorise suggestions for improvements for each arm**: 1. Ran prompt C: we used a script in R to extract all feedback that contains suggestions for improvement, listed them, and categorised them into different types of suggestions with a limit of four categories for the first ~150 responses. 2. Ran prompt C again: we ran the script again on the second ~150 responses. 3. The LLM generated eight or less categories of suggestions for improvement for each arm, totalling approximately 40 categories. 4. We then manually looked at the output to identify similarities and differences in categories across treatment arms. The categories across treatment arms were very similar, so we organised them into four overarching suggestions for improvement across all treatment arms (see page 47).

[1]Note that we used the following large language model (LLM) for all analysis: OpenAI. (2023). *ChatGPT* (Dec 7 version) [Large language model]. https://chat.openai.com/chat

Notes on our use of large language models.[1]

Why did we undertake an AI-assisted approach?
We had a large amount of free text responses (2,164 after cleaning), so we decided to utilise an LLM to assist us in summarising and categorising these responses.

How did we remove personal data?
To remove any personal data from our sample before inputting it into the LLM, we ran a script to exclude the following personal information:

- **Email addresses:** remove any entries containing '@'s
- **Addresses:** remove any entries containing postcodes
- **Phone numbers:** remove any entries containing 9-11 digits
- **Names:** remove any entries which match first names imported from a database of 10,000 names.

We took 10% of the remaining sample and manually checked them. We did not find any personal information in those entries.

Caveats of this approach
LLMs have a number of known limitations, including:

1. **Consistency:** LLMs can provide different outputs to the same prompts. For example, it creates slightly different categories each time you run the prompt.
2. **Reliability** - LLMs do not always exactly follow your prompts and do not always follow instructions correctly. This means that there is a possibility of being misled, as we might think the LLM has followed our instructions based on the outputs it has generated when it actually has not.

This was also the first time BIT have used this approach, and our approach to using LLMs is a work in progress. We therefore remain cautious in our interpretation of the findings and welcome any thoughts or feedback on our approach. Please get in touch at gambling@bi.team.

[1]Note that we used the following large language model (LLM) for all analysis: OpenAI. (2023). *ChatGPT* (Dec 7 version) [Large language model]. https://chat.openai.com/chat

Notes on our use of large language models.[1]

What steps did we take to check the reliability of using LLMs?

Researcher inter-rater reliability check:
For this, two researchers conduct the same measurement or observation on the same sample and their measurements or observations are compared. Our method included:

1. Randomly sampling 75 responses from our overall sample (n =2,164).
2. Two researchers rating the responses as positive, neutral, negative, or n/a. In order to continue with the analysis, a threshold of 80% had to be met using the percentage agreement method.[2] Our percentage agreement was 84% which met our threshold.
3. We then manually combined the responses of both researchers by collaboratively re-examining feedback and selecting one rating as well as changing ‘n/a’s to ‘neutral’ so there were three categories in total.

LLM inter-rater reliability check:

4. We then ran a script in R which asked an LLM to run sentiment analysis on the same sample and conducted an inter-rater reliability check between our combined researcher score and the LLM score. Our percentage agreement was 80%, which met our 80% threshold, so we therefore proceeded to use an LLM. ***Please note we conducted this as an initial check rather than before each time we ran a script which used an LLM.*** *This is because we tested a variety of prompts and scripts across both R and Python and it would be too resource intensive to conduct a reliability each time.*

Categorisation
The categorisation process below was solely conducted in R. For the same 75 responses, we manually conducted a light-touch analysis which involved reading the responses and looking for patterns in the data to extract categories. We found six categories and used these to ‘sense-check’ LLM generated categories from prompt A on page 61. These included: (1) improvements to general design; (2) improvements to the cooling off period; (3) easy to use; (4) efficacy concerns; (5) difficult to understand, and (6) misunderstandings. There was notable crossover between our categories and the LLM’s categories (see page 46), indicating that the LLM output was sufficiently accurate.

[1]Note that we used the following LLM for all analysis: OpenAI. (2023). *ChatGPT* (Dec 7 version) [Large language model]. https://chat.openai.com/chat

[2] O’Connor, C., & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. International Journal of Qualitative Methods, 19, 1609406919899220. https://doi.org/10.1177/1609406919899220

THE
BEHAVIOURAL INSIGHTS TEAM

PREDICTIV

Get in touch:

Project lead
Rachel O'Boyle
rachel.oboyle@bi.team

Project Oversight
Ruth Persian
ruth.persian@bi.team

Eleanor Collerton, Senior Advisor
Deelan Maru, Policy Advisor
Adam Jones, Research Advisor
Steve Human, Research Advisor
Thea House, Associate Research Advisor
Akila Ranganathan, Associate Research Advisor
Louis Shaw, Associate Research Advisor

© Behavioural Insights Ltd.