FRAMEWORK

The SHIELD Framework

SHIELD is a six-stage, evidence-based behaviour change framework built for cybersecurity: Specify, Hypothesise, Intervene, Embed, Learn and Diffuse.

IN THIS DOCUMENT

SHIELD in just over two minutes.	2
Why SHIELD	2
The six stages	3
Models and evidence	8
Running it	10
References	11
Glossary	12

A six-stage behavioural change framework built purely for cybersecurity. Specify, Hypothesise, Intervene, Embed, Learn, Diffuse, with sustainment engineered into every turn of the cycle. Every stage grounded in published evidence, every source cited.

SHIELD in just over two minutes.

Watch the six stages run as one cycle, each shown through the same worked example used on this page: getting a payments team to verify bank detail changes by phone.

2:16

The SHIELD Framework: Specify, Hypothesise, Intervene, Embed, Learn, Diffuse, with one example followed through all six.

Awareness measures attention. SHIELD changes action.

Most security awareness programmes measure attention, not action. People complete training, pass quizzes, and then reuse passwords, click links and share credentials because **the conditions surrounding the behaviour never changed**. The evidence on this is not new: awareness campaigns routinely fail to change behaviour because knowing what to do and doing it are separated by motivation, workload, habit and environment[3], and users make rational trade-offs inside systems that were never designed with their limits in mind[1].

SHIELD is **not a general behaviour change model with security examples bolted on**. Every stage addresses conditions that are distinctive to security behaviour: the behaviour is secondary to the primary task, the benefit is invisible when it works, the cost falls on the individual while the gain falls on the organisation, and the threat adapts. Five principles underpin the framework, and each constrains what a legitimate security behaviour programme is allowed to do.

1

Security is a secondary task

Nobody comes to work to be secure. They come to close deals, treat patients, ship code. Much rejection of security advice is economically rational: the time cost of following it often exceeds the loss it prevents[9]. Employees hold a finite compliance budget, a reserve of effort they will spend on security before they start cutting corners[4]. **Any programme that ignores the budget will exhaust it.**

2

Knowledge does not produce behaviour

The gap between awareness and action is one of the most replicated findings in behavioural science, and security is no exception[3]. Behaviour occurs when capability, opportunity and motivation converge[13], or when motivation, ability and a prompt coincide at the same moment[8]. **Training addresses capability only.** The other components must be engineered.

3

Fear has a ceiling

People act on threat messages when they perceive the threat as real and believe their response will work[17]. When threat is high but efficacy is low, they control the fear rather than the danger, through avoidance, denial or fatalism[22]. Threat information belongs in a behaviour programme only when **paired with a clear, achievable, effective action**.

4

Friction beats persuasion

Where the environment can be changed, change the environment[20]. **Making the secure path the default path** will outperform any poster campaign, because it removes the need for a decision at all. Applied ethically, choice architecture is a legitimate and well-mapped security lever[15].

5

Unmeasured behaviour regresses

Completion rates and click rates are proxies at best. Security fatigue is a measurable state that degrades decision quality[18], and where official security is unworkable, employees invent shadow security, workarounds invisible to the metrics[10]. Measurement must **triangulate** observed behaviour, validated self-report[7] and the psychological drivers behind both.

Six stages, one operating cycle

SHIELD runs as a cycle, not a project. Nothing is designed until something is diagnosed, nothing is claimed until something is measured, and nothing is scaled until something is validated. Each stage below is paired with a running worked example that follows one behaviour, verifying bank detail changes by phone, from definition through to organisation-wide rollout. Select a stage to explore it.

S

Specify

Define the target behaviours, precisely, and prioritise them by risk

A behaviour programme fails at conception when its target is an abstraction. “Improve security culture” and “raise awareness” are not behaviours. A behaviour is **observable, specific and situated**: *reports a suspicious email within four hours of receipt*, *verifies a payment change request by phone before actioning it*, *locks the workstation when leaving the desk*.

Start from the organisation's actual threat picture, not a generic list. The behaviours that matter in a hospital differ from those in a software house. Established taxonomies of end-user security behaviour help separate behaviours requiring motivation work from those requiring capability work[19], and the SeBIS domains of device securement, password generation, proactive awareness and updating offer a validated structure for the personal hygiene layer[7].

For each candidate behaviour, record the risk it addresses, the population that performs it, and the moment in the workflow where it occurs. Then **prioritise ruthlessly**. A programme targeting three behaviours well will outperform one targeting thirty behaviours nominally, because the compliance budget is spent **per person, not per policy**. [4].

OUTPUT OF THIS STAGE

A ranked register of no more than five target behaviours per audience segment, each written as an observable action with a defined trigger moment.

WORKED EXAMPLE

Stopping payment diversion fraud

Fraud intelligence and two near-misses put payment diversion at the top of the risk picture, so one register entry reads: *any request to change supplier or payroll bank details is verified by phone, using the number held on the master supplier record, before the change is actioned*.

> Population: 14 people across accounts payable and payroll. Trigger moment: a change request landing in the shared inbox or the ERP queue. Vague alternatives like “be alert to payment fraud” were rejected because they cannot be observed or measured.

H

Hypothesise

Diagnose why the behaviour is or is not happening, as a testable hypothesis

The diagnostic instrument is COM-B: behaviour occurs when Capability, Opportunity and Motivation are all sufficient[13]. **A deficit in any one is enough to stop the behaviour**, and each deficit demands a different remedy. The diagnosis is written as a **hypothesis, not a verdict**, because the Learn stage will test it and the cycle depends on being able to say the diagnosis was wrong.

The social opportunity component deserves particular weight in security. Local team norms, **what colleagues actually do rather than what policy says**, are the strongest determinant of real practice[10]. Social proof explains the mechanism: people calibrate against observed peers, especially under uncertainty, and security decisions are almost always made under uncertainty[5].

Gather the evidence through mixed methods: a structured COM-B questionnaire for breadth, short interviews or team observation for depth, and existing telemetry as a cross-check rather than a source of truth. Where self-report and observed behaviour diverge, **that divergence is itself diagnostic**. It usually points at social desirability pressure or at shadow security in operation.

> **OUTPUT OF THIS STAGE**
>
> A per-behaviour diagnostic hypothesis stating which COM-B components are deficient, with evidence, for each audience segment.

> **WORKED EXAMPLE**
>
> ### Stopping payment diversion fraud
>
> Interviews and a short COM-B questionnaire across the 14 staff show: **capability** adequate, everyone can describe the fraud pattern. **Physical opportunity** weak, there is no verified call-back directory and the ERP has no pause point, so verifying means improvising. **Social opportunity** weak, the team is praised for same-day processing and nobody has ever seen a colleague make a verification call. **Motivation** moderate reflectively, absent automatically.
>
> Hypothesis: the behaviour fails on opportunity, not knowledge. More training will not move it.

I

Intervene

Design interventions that match the diagnosis, then screen them with APEASE

The Behaviour Change Wheel links each COM-B deficit to intervention functions that can plausibly correct it[13]. The security translation is set out below.

DIAGNOSED DEFICIT	MATCHED INTERVENTION FUNCTIONS	CYBERSECURITY EXAMPLES
Psychological capability	Education, training, enablement	Scenario-based practice on real lures received by the organisation; just-in-time guidance at the point of decision
Physical opportunity	Environmental restructuring, enablement	One-click reporting in the mail client; password manager deployed, pre-configured and paid for; secure defaults on collaboration tools

DIAGNOSED DEFICIT	MATCHED INTERVENTION FUNCTIONS	CYBERSECURITY EXAMPLES
Social opportunity	Modelling, enablement	Security Champions embedded in teams; leaders visibly reporting their own near-misses; team-level norms made visible
Reflective motivation	Persuasion, incentivisation, education	Feedback loops that show reporters what their report prevented; recognition for protective behaviour rather than punishment for error
Automatic motivation	Habit formation, prompts, environmental restructuring	Consistent cue-action pairing; prompts placed at the trigger moment identified in Specify

Candidate interventions are then screened against the **APEASE criteria** before anything is committed: Affordability, Practicability, Effectiveness and cost-effectiveness, Acceptability, Side-effects and safety, and Equity[14]. APEASE is what stops a theoretically sound intervention that the organisation cannot afford, that the team will resent, or that quietly disadvantages one group, from reaching delivery.

Three design rules apply on top of the mapping. First, **reduce cost before adding motivation**: increasing ability is almost always cheaper and more reliable than increasing motivation[8]. If reporting a phish takes six clicks and a form, no campaign will fix the reporting rate. Make it one click, then campaign.

Second, **design for habit, not decision**. Behaviours that recur at a stable cue should be built as habits: consistent context, immediate confirmation, and repetition until automaticity develops, which takes a median of **66 days** and varies widely between individuals[12]. Habit change is most achievable at moments of context disruption such as onboarding, office moves and new systems, so time interventions to these windows[21].

Third, **keep fear appeals honest**. If a threat message is used, it must carry response efficacy and self-efficacy in the same breath[22]. “Criminals are targeting our payment team” is incomplete and harmful on its own. “Criminals are targeting our payment team; a thirty-second call-back on any bank detail change defeats them, and here is the number” is a legitimate intervention.

OUTPUT OF THIS STAGE

An APEASE-screened intervention specification per behaviour, stating the deficit it corrects, the mechanism it uses, the sources it draws on, and the moment it will be delivered.

WORKED EXAMPLE

Stopping payment diversion fraud

The package matched to the opportunity deficit: a mandatory hold state in the ERP for any bank detail change (environmental restructuring), a verified call-back directory owned by procurement (enablement), and the team lead performing the first call-backs in the open-plan office (modelling).

Two rejected candidates show APEASE working. A poster campaign failed on effectiveness, since the diagnosis was not a knowledge gap. Extra e-learning failed on acceptability for a team already at capacity, and would have spent compliance budget for nothing.

E

Embed

Deliver through the workflow and through people, not through a channel called awareness

Delivery lives or dies on **proximity to the trigger moment**. An instruction delivered in an annual module is gone within days; the same instruction delivered as a prompt at the moment of decision has a fighting chance[8]. Wherever technically possible, interventions ship inside the tools where the behaviour happens: the mail client, the code repository, the payment system.

The second delivery mechanism is people. Security Champions, respected peers inside teams who model, translate and advocate for secure behaviour, exploit the social proof and local norm effects surfaced in Hypothesise. Champion networks are one of the few practices that reliably distinguish organisations that changed behaviour from those that merely trained[2]. Champions collapse the distance between “the security team says” and “we do it this way here”. They need selection for **credibility rather than seniority**, genuine time allocation, direct lines back into the security function, and recognition. A Champions network run as an unfunded mailing list is theatre.

Delivery must also protect the compliance budget [4]. Sequence interventions so that no audience receives more than **one behavioural ask at a time**, retire asks that measurement shows are now habitual, and publicly remove security requirements that Hypothesise exposed as pointless. **Removing a useless control is itself an intervention:** it buys back budget and signals that the security function is paying attention.

> **OUTPUT OF THIS STAGE**
>
> Interventions live in the workflow, a resourced Champions network, and a delivery calendar that enforces one ask at a time per audience.

> **WORKED EXAMPLE**
>
> ### Stopping payment diversion fraud
>
> The hold state ships inside a scheduled ERP release. When a change request is opened, a one-line prompt appears at the decision point: *verify by phone before release, directory here*, with the directory one click away.
>
> The finance champion, a senior AP clerk the team already turns to, sits with colleagues through the first week of call-backs. Nothing else is asked of this audience all quarter, and a legacy monthly attestation form the diagnosis exposed as pointless is publicly retired, buying back the budget the new behaviour spends.

L

Learn

Measure behaviour and its drivers, not exposure to content

The measurement layer triangulates three sources, and **the triangulation is the point**. **Observed behaviour:** reporting rates and time-to-report, verification call volumes, screen-lock compliance, credential hygiene signals, and simulation outcomes where simulations are run ethically, with no entrapment lures, no naming and shaming, and feedback in the moment. Observed data answers “what changed”. **Self-reported behaviour and intention:** validated instruments such as SeBIS for the hygiene domains[7], supplemented by items built for the organisation's specific target behaviours. Self-report answers “what do people believe they do”, and its divergence from observation is a finding, not a nuisance. **Drivers:** the COM-B diagnostic repeated at intervals to detect whether capability, opportunity and motivation are actually moving, with security fatigue items included[18], because a programme can improve behaviour in quarter one by methods that exhaust people by quarter three.

Set baselines during Hypothesise, define success thresholds during Intervene, and review on a fixed cadence. Treat every intervention as an experiment with a **falsifiable prediction**. Where the prediction fails, the hypothesis was wrong, and the cycle returns to Hypothesise rather than shouting the same message louder. Where it holds, the intervention earns its place in Diffuse.

One warning belongs here explicitly. Do not collapse this measurement layer into an individual risk score derived from telemetry alone. A score built purely from clicks and completions measures **exposure and logging coverage, not the person**, and steering interventions by it invites both unfairness and gaming. Scores, where used at all, should sit at team or process level and always alongside the driver data that explains them.

OUTPUT OF THIS STAGE

A behavioural dashboard per target behaviour showing observed change, self-reported change, driver movement and fatigue, against baseline and threshold, with a verdict on each hypothesis.

WORKED EXAMPLE

Stopping payment diversion fraud

Prediction set before launch: within eight weeks, verification calls logged will match bank detail change requests at 95 per cent or better, with no rise in team fatigue scores.

Week eight: 97 per cent match. Two genuine diversion attempts caught, one worth £48,000, both fed back to the team within days. Self-report and call logs agree, fatigue is flat. Hypothesis confirmed, and the package qualifies for Diffuse. Had the match rate stalled at 60 per cent, the team would have returned to Hypothesise, not commissioned more comms.

D

Diffuse

Scale only what Learn validated, and adapt it to each context it enters

A pilot that worked in one business unit is **evidence, not victory**. Diffuse is the disciplined spread of validated interventions across the organisation, and it has its own science. Innovations spread when adopters can see relative advantage, when the change is compatible with how they already work, when it is simple to try, and when its results are observable[16]. An intervention that succeeded in the pilot because a respected team lead backed it will stall in a unit where nobody plays that role, so each wave of rollout begins with a light re-run of Hypothesise for the receiving context.

The Champions network built during Embed is the natural diffusion infrastructure. Champions function as the opinion leaders and early adopters through whom new practice reaches the majority[16, 2], which means the network map matters: a wave should land where champion coverage is strong, and gaps in coverage are addressed before rollout rather than discovered during it.

Two rules keep scaling honest. First, **adapt the periphery, protect the core**: local units may change the wording, the channel and the timing, but the behavioural mechanism the pilot validated is not negotiable, otherwise the evidence no longer applies. Second, meter the pace against the compliance budget[4]: **a wave per audience at a time**, with Learn measurement running at each wave so that decay or local failure is caught early rather than averaged away in an organisation-wide number.

OUTPUT OF THIS STAGE

A wave-based rollout plan for each validated intervention, with per-context diagnosis, champion coverage confirmed, and measurement gates between waves.

WORKED EXAMPLE

Stopping payment diversion fraud

Wave one takes the package to the two largest business unit finance teams, both with strong champion coverage. One runs a different ERP, so its hold state is rebuilt natively rather than copied; the mechanism, a forced pause plus a verified directory, is preserved exactly.

A third unit has no champion, so one is recruited and embedded before its wave begins rather than after it fails. Wave two waits behind a measurement gate: it starts only when wave one repeats the 95 per cent threshold.

The Sustainment Layer

Sustainment is not a stage, it is a property of the running cycle. Initiation and maintenance are different problems with different mechanisms: changed behaviour persists through sustained motives, self-regulation skill, habit, a supportive environment and social support[11]. SHIELD engineers those conditions continuously across every stage, and the loop itself closes the register: Learn and Diffuse feed Specify on an annual cycle or after any material incident, retiring behaviours that have become habitual and promoting new ones the threat picture demands.

Feedback keeps motives alive

People who report phishing and hear nothing stop reporting. Every protective act gets a **closed loop within days**: what the report or call-back was, and what happened because of it. “Your call-back stopped a £48,000 diversion attempt” sustains a behaviour better than any campaign.

Just culture keeps people in the system

Lapse recovery is made **cheap and blame-free**. A mistaken click met with help rather than sanction keeps people inside the system instead of hiding from it[10, 1]. This is psychological safety applied to security, and it keeps the reporting channel honest.

Habits are re-anchored at disruption

Habit survives on environmental stability. Every system rollout, restructure and office move is a scheduled re-run of Hypothesise, Intervene and Embed for the behaviours it touches, because **disruption is when habits break** and when new ones form most easily[21].

Culture is the output, not the input

Information security culture is the accumulated result of what the organisation repeatedly does, rewards and tolerates[6]. **It cannot be declared, only earned** by running the cycle consistently. Leadership behaviour is the loudest signal: an executive who verifies a payment request by phone in front of their team does more than any campaign asset.

The models and evidence SHIELD is built on

SHIELD is a synthesis, and it is open about its parts. Each model, theory or instrument below is used at a specific point in the cycle, marked by its stage letter. The Sustainment marker denotes the Sustainment Layer.

COM-B Model

H · L

The diagnostic core: behaviour requires Capability, Opportunity and Motivation, and each deficit demands a different remedy[13].

Behaviour Change Wheel

I

Maps each diagnosed COM-B deficit to the intervention functions that can plausibly correct it[13].

APEASE criteria

I

Screens candidate interventions for Affordability, Practicability, Effectiveness, Acceptability, Side-effects and Equity before commitment[14].

Fogg Behaviour Model

I · E

Behaviour happens when motivation, ability and a prompt converge; drives prompt placement at the trigger moment and ability-first design[8].

Protection Motivation Theory

I

Explains when threat messaging works: perceived severity and vulnerability paired with response efficacy and self-efficacy[17].

Extended Parallel Process Model

I

Explains when fear appeals backfire: high threat with low efficacy produces fear control, not danger control[22].

Compliance Budget

S · E · D

Employees hold a finite effort reserve for security; SHIELD paces asks, retires dead controls and meters rollout against it[4].

Nudge and choice architecture

I

Secure defaults and reduced friction change behaviour without persuasion, applied within published ethical guardrails[20, 15].

Habit formation research

I · E

Automaticity develops through consistent cue-action pairing over a median of 66 days, and context disruption is the window for change[12, 21].

Social proof and influence

H · E

People calibrate against observed peers under uncertainty; the mechanism behind local norms and the Champions model[5].

Diffusion of Innovations

D

Validated practice spreads through opinion leaders when advantage is visible, adoption is compatible and results are observable[16].

SeBIS

S · L

A validated Security Behavior Intentions Scale providing structured self-report across the personal hygiene domains[7].

End-user behaviour taxonomy

S

Separates behaviours needing motivation work from those needing capability work when building the register[19].

Security fatigue

L

A measurable exhaustion state that degrades security decisions; tracked so gains in one quarter do not exhaust people in the next[18].

Shadow security

H

Workarounds reveal where official security is unworkable; treated as diagnostic signal rather than disobedience[10].

Behaviour maintenance theory

Sustainment

Changed behaviour persists through sustained motives, self-regulation, habit, environment and social support[11].

Security culture research

Sustainment

Culture as the accumulated output of repeated organisational practice, assessable and earned rather than declared[6].

Running the framework

The single most common implementation failure is entering at Intervene, buying an intervention because a vendor sold it convincingly, then reverse-fitting a justification. The second is scale without validation. Five behaviours, properly diagnosed, precisely intervened upon, honestly measured and then diffused, will shift

real risk. Fifty behaviours addressed through content distribution will shift completion statistics. A realistic operating rhythm for a mid-sized organisation looks like this.

Specify + Hypothesise

Six to eight weeks of behaviour definition, prioritisation and diagnosis per cycle.

Intervene

Four weeks of APEASE-screened design matched to the diagnosis, with predictions set.

Embed

Continuous delivery through workflow, prompts and the Champions network.

Learn

Quarterly review of observed behaviour, self-report, drivers and fatigue against baseline.

Diffuse

Wave-based scaling of validated interventions, with measurement gates between waves.

Sustainment Layer

Standing practice across all stages, with a full-cycle refresh annually or after any material incident.

References

SHIELD synthesises published behavioural science and security usability research. Everything drawn from others is cited below.

1. Adams, A. and Sasse, M.A. (1999) 'Users are not the enemy', *Communications of the ACM*, 42(12), pp. 40–46.
2. Alshaikh, M. (2020) 'Developing cybersecurity culture to influence employee behavior: A practice perspective', *Computers & Security*, 98, 102003.
3. Bada, M., Sasse, M.A. and Nurse, J.R.C. (2015) 'Cyber security awareness campaigns: Why do they fail to change behaviour?', *Proceedings of the International Conference on Cyber Security for Sustainable Society*, pp. 118–131.
4. Beautement, A., Sasse, M.A. and Wonham, M. (2008) 'The compliance budget: Managing security behaviour in organisations', *Proceedings of the New Security Paradigms Workshop (NSPW)*, pp. 47–58.
5. Cialdini, R.B. (2007) *Influence: The Psychology of Persuasion*. Revised edn. New York: Harper Business.
6. Da Veiga, A. and Eloff, J.H.P. (2010) 'A framework and assessment instrument for information security culture', *Computers & Security*, 29(2), pp. 196–207.
7. Egelman, S. and Peer, E. (2015) 'Scaling the security wall: Developing a Security Behavior Intentions Scale (SeBIS)', *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 2873–2882.
8. Fogg, B.J. (2009) 'A behavior model for persuasive design', *Proceedings of the 4th International Conference on Persuasive Technology*, Article 40.
9. Herley, C. (2009) 'So long, and no thanks for the externalities: The rational rejection of security advice by users', *Proceedings of the New Security Paradigms Workshop (NSPW)*, pp. 133–144.

10. Kirlappos, I., Parkin, S. and Sasse, M.A. (2014) 'Learning from “shadow security”: Why understanding non-compliant behaviors provides the basis for effective security', *Proceedings of the Workshop on Usable Security (USEC)*.
11. Kwasnicka, D., Dombrowski, S.U., White, M. and Sniehotta, F. (2016) 'Theoretical explanations for maintenance of behaviour change: A systematic review of behaviour theories', *Health Psychology Review*, 10(3), pp. 277–296.
12. Lally, P., van Jaarsveld, C.H.M., Potts, H.W.W. and Wardle, J. (2010) 'How are habits formed: Modelling habit formation in the real world', *European Journal of Social Psychology*, 40(6), pp. 998–1009.
13. Michie, S., van Stralen, M.M. and West, R. (2011) 'The behaviour change wheel: A new method for characterising and designing behaviour change interventions', *Implementation Science*, 6, 42.
14. Michie, S., Atkins, L. and West, R. (2014) *The Behaviour Change Wheel: A Guide to Designing Interventions*. London: Silverback Publishing.
15. Renaud, K. and Zimmermann, V. (2018) 'Ethical guidelines for nudging in information security & privacy', *International Journal of Human-Computer Studies*, 120, pp. 22–35.
16. Rogers, E.M. (2003) *Diffusion of Innovations*. 5th edn. New York: Free Press.
17. Rogers, R.W. (1975) 'A protection motivation theory of fear appeals and attitude change', *The Journal of Psychology*, 91(1), pp. 93–114.
18. Stanton, B., Theofanos, M.F., Prettyman, S.S. and Furman, S. (2016) 'Security fatigue', *IT Professional*, 18(5), pp. 26–32.
19. Stanton, J.M., Stam, K.R., Mastrangelo, P. and Jolton, J. (2005) 'Analysis of end user security behaviors', *Computers & Security*, 24(2), pp. 124–133.
20. Thaler, R.H. and Sunstein, C.R. (2008) *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven: Yale University Press.
21. Verplanken, B. and Wood, W. (2006) 'Interventions to break and create consumer habits', *Journal of Public Policy & Marketing*, 25(1), pp. 90–103.
22. Witte, K. (1992) 'Putting the fear back into fear appeals: The extended parallel process model', *Communication Monographs*, 59(4), pp. 329–349.

Glossary

The terms from network and behavioural science used in this document, in plain English.

Automaticity. The point at which a behaviour runs on cue without conscious thought, built by repeating the same action in the same context.

Capability, opportunity and motivation. A behavioural diagnosis model: behaviour happens only when Capability (can they do it), Opportunity (does the environment allow and encourage it) and Motivation (do they want to, consciously or by habit) are all present at once.

Choice architecture. Designing environments and defaults so the secure option is the easiest one, removing the need for a conscious decision at all.

COM-B. A behavioural diagnosis model: behaviour happens only when Capability (can they do it), Opportunity (does the environment allow and encourage it) and Motivation (do they want to, consciously or by habit) are all present at once.

Compliance budget. The finite reserve of time and effort employees will spend on security before they start cutting corners. Every security ask draws it down, and it does not refill by asking harder.

Opinion leaders. The trusted individuals a group watches when deciding whether to adopt something new. Influence follows credibility, not the org chart.

Psychological safety. The shared belief that it is safe to speak up, admit mistakes and ask questions without fear of punishment or embarrassment. Without it, incidents go unreported.

Response efficacy. Response efficacy: the belief that the recommended action actually works against the threat. Without it, threat messages produce anxiety rather than action.

Security fatigue. A measurable weariness caused by too many security demands and decisions, which degrades judgement and drives risky shortcuts.

Self-efficacy. Self-efficacy: the belief that you personally are capable of performing the recommended action in your real working conditions.

Shadow security. The unofficial workarounds people invent when official security gets in the way of the job. Invisible to compliance metrics, but it defines how work really gets done.

Social proof. The human tendency to copy what visible peers do, especially under uncertainty. If nobody around you verifies requests, you are unlikely to start.

Telemetry. Data collected automatically by systems: clicks, logins, completions, log events. It records what happened, never why.

About this document

This document was built on 24 November 2026 from the page at cybehave.com/shield-framework, so it carries the guidance as it was published that day. The page is kept up to date; download a fresh copy whenever you need one. On the page, the Intervention Designer walks one behaviour through the SHIELD method in your browser and exports the plan.

This page **cybehave.com/shield-framework**

Free Champions Programme Health Check **cybehave.com/champions-health-check/**

The Rise of the Security Champion, free ebook **cybehave.com/the-rise-of-the-security-champion**

HEROES by CyBehave, free 28-day trial **cybehave.com/trial**

© 2026 CyBehave Ltd. You are welcome to share this document with colleagues. Please keep it whole and credit CyBehave.