

VII

Zrychlí ti stroj
i chaos?

Po této kapitole víš, že nástroj nezná směr: zesílí dobrý proces i zmatek stejně ochotně. Umiš spočítat, kdy se lavina rozjede a kdy vyhasne — je to tatáž matematika jako u epidemie — a víš, jak si postavit protokol, aby ti AI nevyráběla jen delší, zdvořilejší chaos.

První pravidlo každé technologie nasazené v provozu zní, že automatizace aplikovaná na efektivní operaci efektivitu znásobí. Druhé zní, že automatizace aplikovaná na neefektivní operaci znásobí neefektivitu.

— Bill Gates, *The Road Ahead* (1995), kap. 8. Napsáno mužem, v jehož firmě se o dva roky později odehrál *Bedlam DL3*.

Bedlam DL3

Čtrnáctého října 1997 si jeden zaměstnanec Microsoftu — tehdy největší softwarové firmy světa, plné nejlepších inženýrů, jaké si mohla koupit — všiml, že je v jakémsi distribučním seznamu jménem *Bedlam DL3*, ke kterému nemá důvod patřit. Udělal jedinou rozumnou věc, jakou v takové chvíli člověk udělá: napsal e-mail s prosbou „vyřadte mě odsud“. Jenže *Bedlam DL3* nebyl malý seznam kolegů z patra. Byl to jeden z obřích seznamů, které IT oddělení nastavilo kdysi ke správě serverů — a byly na něm *desetitisíce* adres. Podle blogu exchangeového týmu kolem pětadvaceti tisíc; jeden pamětník mluví o třinácti tisících, čtvrtině všech firemních schránek. Přesné číslo si nepamatuje nikdo, protože za chvíli už nikdo nepočítal.

Ta prosba „vyřadte mě“ totiž nedorazila správci seznamu. Dorazila všem. A tady začíná lavina, kterou znáš, i kdybys o *Bedlamu* nikdy neslyšel, protože ji zažil každý, kdo kdy pracoval ve větší firmě. Desítky lidí odpověděly „mě taky vyřadte“ — a protože tlačítko *Odpovědět všem* je hned vedle, odpověděly všem. Na to začali další psát „přestaňte prosím odpovídat všem!“ — a poslali to, pochopitelně, všem. Přišly vtipy. Přišly rozhořčené výzvy k

rozvaze. Přišly odpovědi na vtipy a rozhořčené odpovědi na výzvy k rozvaze. Každá jednotlivá zpráva byla lidsky pochopitelná; dohromady tvořily bouři.

A byla to bouře v doslovném, měřitelném smyslu. Za necelou hodinu vzniklo kolem *patnácti milionů* zpráv, sítě protéklo asi sto pětadevadesát gigabajtů dat a poštovní servery firmy, která ten poštovní software vyráběla a prodávala celému světu, pod tou vahou lehly. Uklízelo se ještě další dva dny. Nešlo o žádný útok, žádnou chybu v kódu, žádného hackera. Systém fungoval *přesně tak, jak měl*: doručoval poštu. Jen jí bylo najednou patnáct milionů kusů, protože každá odpověď plodila další odpovědi rychleji, než je stačil kdokoli číst.

A teď ta pointa, kvůli které je Bedlam v knize o vibe codingu. Nezhroutili ho hlupáci. Zhroutili ho chytří, dobře míněně jednající lidé, z nichž každý udělal drobnou, lidsky srozumitelnou chybu — a mezi ně postavili nástroj, který tu chybu dokázal zesílit a rozeslat rychleji, než ji kdokoli stačil zastavit. Nikdo nechtěl bouři. Všichni chtěli klid. Bouře přišla proto, že systém byl *zesilovač* — a zesilovač nezná směr.

Reply-All Storm Protection přišel do Exchange až dodatečně: nástroj pro správce zastavit bouři kolem roku 2008, automatická detekce a zadušení až v roce 2020 — třináct let po Bedlamu. Ochrana proti lavině není v systému zdarma; musí se dostavět schválně.



Nástroj bez směru

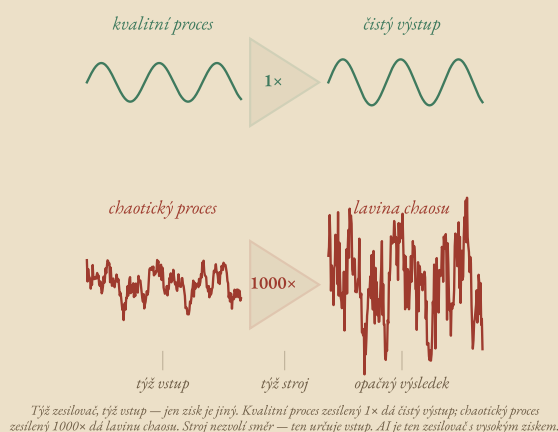
Nezačínáme u Microsoftu. Začneme u sebe, protože jinak by z téhle kapitoly byl posměšek nad cizí neschopností — a to je přesně to, čím být nechce.

Otevři si schránku ve firmě, kde jsem pracoval, a najdeš tam celou galerii hrůzy — a moje jméno je pod pořádným kusem z ní. Řetězy s dvaceti lidmi v kopii, z nichž devatenáct řešit nemuselo nic. E-mail, kterým jsem „všem“ vysvětloval, ať už laskavě přestanou psát „všem“ — poslaný, jak jinak, všem. Ukřivdění vlákno, které jsem rozjel v přesvědčení, že mám pravdu, a které se po třech dnech rozpustilo v třiceti odpovědích, jež nikam nevedly. Padesát let existuje e-mail. Padesát let. A my — já — ho pořád neumíme použít tak, aby po nás nezůstala spoušť. To není historka o někom hloupém. To jsem já a nejspíš i ty.

Přiznávám to hned na začátku schválně, protože jinak by nám ta pointa utekla. Snadno se řekne „no jistě, ti lidé v Microsoftu neuměli e-mail“. Jenže e-mail neumíme *my všichni* — po půl století praxe. Ne proto, že jsme hlupí, ale proto, že e-mail je nástroj, který stejně ochotně doručí promyšlenou zprávu i unáhlenou, jednu adresu i dvacet tisíc. Nezná rozdíl mezi tím, co má cenu poslat, a tím, co nemá. Jen doručuje. A my do něj nasypeme svůj zmatek a on ho věrně roznese dál.

A tady je most k celé téhle knize, protože přesně tenhle nástroj teď dostal mladšího, nesrovnatelně silnějšího bratra. Jazykový model je taky zesilovač — jen s mnohem vyšším ziskem. Když do něj vejdeš s jasnou hlavou, s dobře položenou otázkou a s procesem, jak výstup ověřit, znásobí tvou práci způsobem, který ještě před pár lety nešel. Ale když do něj vejdeš se zmatkem — s nejasným zadáním, s dojmem místo specifikace, s „nějak to udělej“ — nedostaneš z něj jasno. Dostaneš *delší, zdvořilejší a početnější verzi téhož zmatku*. Vygeneruje ti pět odstavců tam, kde jsi měl v hlavě kaši; napíše ti tři varianty řešení problému, který jsi špatně pochopil; rozešle tvůj nedomyšlený nápad hladce a přesvědčivě dvaceti lidem. Chaos nezmizel. Jen se zrychlil a získal lepší sloh.

Zesilovač nezná směr



Zesilovač nezná směr. Nosný obraz kapitoly: stroj zesílí dobrý proces i chaos, bez rozdílu. Vrací se v kap. 16, kde model odměněný za potlesk zesiluje i tvůj vlastní omyl — podlézavě a sebejistě.

Podívej se na obrázek, protože je to celá teze kapitoly v jednom schématu. Nahoře i dole je *tentýž* zesilovač a do obou vchází *tentýž* druh signálu. Nahoře je vstup čistý a zisk mírný: co vejde, to vyjde zesílené a pořád čisté. Dole je vstup roztřepený, plný šumu, a zisk je tisícinásobný — a co vyjde, není o nic čistší, jen o tisíc řádů hlasitější. Stroj nezvolí, který scénář nastane. Směr určil vstup. Zesilovač jen dodal energii. A to je přesná definice toho, co dnes držíš v ruce: obrovský zisk připojený k tomu, co do něj vložíš. Jestli je to jasno, dostaneš zesílené jasno. Jestli je to zmatek, dostaneš zesílený zmatek — a rychleji, než ho stihneš přechytit.

Poctivě řečeno, „napiš to za mě“ je nejmilejší věta desetiletí a sám ji říkám denně. Jenom mi u ní nikdo neřekl, že když nevím, co chci napsat, dostanu to nevědění zpátky — jen o tři odstavce delší, se závěrečným shrnutím a nabídkou, že to klidně rozvede ještě víc.

To je také důvod, proč tahle kniha trvá na tom, že metoda musí předcházet nástroji, a ne naopak. Kdo umí e-mail — ví, komu píše, proč a co po nich chce —, tomu rychlejší doručování pomůže. Kdo e-mail

neumí, tomu rychlejší doručování jen rychleji rozešle spoušt. Totéž platí o modelu do puntíku. Nástroj nenahrazuje proces; zesiluje ten, který už máš. Když žádný nemáš, zesílí prázdno — a prázdno zesílené na tisícínásobek je pořád prázdno, jen zabere víc místa.

Protokoly, které škálují

Když je problém v tom, že zesilovač zesílí zmatek, řešení nezní „zeslab zesilovač“ — to by bylo jako přestat používat e-mail nebo vypnout model. Řešení zní: *dej na vstup pořádek*. A pořádek, který vydrží i pod zesílením, se nedělá dobrou vůlí ani snahou být pečlivý. Dělá se *protokolem* — pár tvrdými pravidly, která rozhodnou předem, ještě než se v reálném čase začne dít chaos. (Že rozhodnutí učiněné předem poráží dobrou vůlí v reálném čase, je celá příští kapitola; tady je jen první ochutnávka.)

Nejlepší doklad, jak mocný umí být triviálně jednoduchý protokol, přinesl chirurg Atul Gawande. Vzal nemocnice — prostředí plné špičkových lékařů, tedy zase ti nejlepší, jako v Microsoftu — a zavedl na operačních sálech *checklist*. Ne školení, ne motivační přednášku, ne chytřejší lékaře. Odškrtnávací seznam banalit typu „potvrdili jsme jméno pacienta a stranu, kterou operujeme“. Komplikace a úmrtnost po zavedení ztlačily. Ne proto, že by chirurgové najednou zmoudřeli, ale proto, že checklist odebírá rozhodování z místa, kde je člověk pod tlakem a chybuje, a přesouvá ho do klidu předem. Slabý článek není znalost. Slabý článek je spolehnout se v reálném čase na to, že si vzpomeneš.

Protokol pro výstupy AI. Než pošleš dál cokoli, co model vygeneroval, rozsekej to na tři druhy vět — protože každý druh se ověřuje jinak.

KDO JE PŘÍJEMCE?

Člověk? stroj? dvacet lidí v kopii? produkční systém?

Čím dál doletíš, tím dražší je chyba → tím tvrdší protokol.

1) OVĚŘITELNÉ TVRZENÍ ("funkce vrací setříděný seznam")

Musí jít nezávisle zkontrolovat. Napiš JAK to ověříš, než to pošleš dál — jinak posíláš víru, ne fakt.

2) AKCE ("smaž tabulku", "odešli všem", "nasad verzi")

Nevratné akce potřebují potvrzení a undo PŘEDEM. Kolik věcí se stane, když to spustíš? (→ lavina)

3) OZDOBA (zdvořilosti, úvod, shrnutí, "rád pomůžu")

Nenese informaci ani akci. Model jí vyrábí nejvíc.

Nekontroluj ji — smaž ji.

Všimni si, co ten protokol dělá, protože v tom je celý trik. Nesnaží se model zpomalit ani ztlumit. Nechává zesilovač zesilovat naplno — a jen *pořádá vstup* tak, aby to, co se zesílí, stálo za zesílení. Rozdělíš, co model vyprodukoval, na tvrzení (ta si ověříš), akce (ty pojistíš proti nevratnosti) a ozdobu (tu zahodíš). Rázem není otázka „věřím tomu, co model napsal?“, na kterou nikdo neumí odpovědět, ale tři menší otázky, na které odpověď je: čím to ověřím, co se stane při spuštění, a co je vata. To je adresát rovná se akce z nejlepší e-mailové etikety, přeložené do práce se strojem: každá věta ať ví, jestli je fakt k ověření, čin k pojištění, nebo vzduch k vymazání.

A tady je ta věta, kterou si z celé dílny odnes, protože platí i pro e-mail, i pro model, i pro cokoli dalšího, co ti kdy dají do ruky jako zesilovač: „*napiš to za mě*“ *bez struktury nevyrábí pořádek, vyrábí delší chaos*. Model neumí dodat proces, který nemáš — umí jen zesílit ten, který přineseš. Když přineseš protokol, zesílí pořádek. Když přineseš „nějak to zaříd“, zesílí „nějak“. Struktura na vstupu není byrokracie navíc. Je to jediné místo, kde ještě máš kontrolu nad tím, co ze zesilovače vyleze.

Gawande, The Checklist Manifesto (2009): pilotní studie WHO na osmi nemocnicích ukázala pokles komplikací i úmrtnosti po zavedení chirurgického checklistu. Lekce není „lékaři jsou nedbalí“ — je „paměť pod tlakem selhává i u expertů; protokol ji obejde“.

Věť: kdy se lavina rozjede

Zvedneme to o patro, protože pod tou vtipnou historkou o e-mailech se schovává jedna z nejhezčích a nejpřenositelnějších rovnic v celé knize — a platí úplně stejně pro reply-all bouři, pro epidemii i pro atomovou bombu. Nebude to fingovaná matematika. Bude to jeden pojem, jedna rovnice a jeden práh, který rozhoduje o všem.

Ptejme se úplně přímočaře: kdy se z jedné zprávy stane bouře patnácti milionů, a kdy naopak zapadne bez odezvy? Klíč je v jediném čísle. Nazvěme ho R : *očekávaný počet odpovědí, které vyvolá jedna zpráva*. Když napíšeš do velkého seznamu, kolik lidí v průměru odpoví? Když jedna odpověď vyvolá v průměru R dalších, a každá z nich zase R dalších, pak počet zpráv v jednotlivých vlnách jde

$$N_0 = 1, \quad N_1 = R, \quad N_2 = R^2, \quad \dots, \quad N_n = R^n.$$

To je celý mechanismus. Jedna úvodní zpráva ($N_0 = 1$), první vlna odpovědí ($N_1 = R$), druhá vlna ($N_2 = R^2$) a tak dál geometrickou řadou. A teď to podstatné: chování téhle řady visí *celé* na tom jediném čísle R — na tom, jestli je menší, rovno, nebo větší než jedna.



EINSTEIN · MATEMATIKA — přeskočitelné

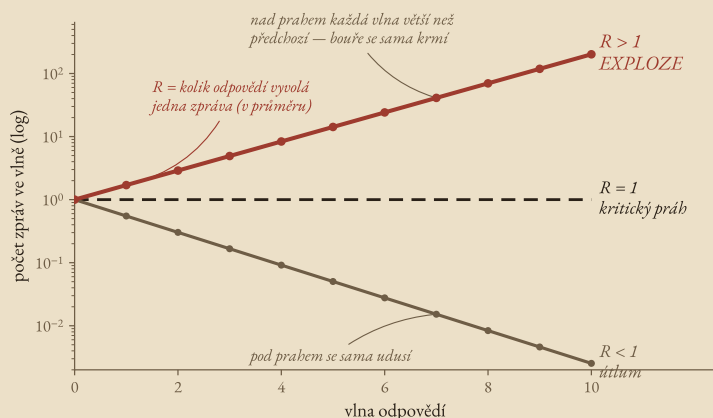
Větvící se proces a kritický práh. Označ R střední počet potomků, které jedna zpráva vyvolá. Počet zpráv v n -té vlně je $N_n = R^n$ a celkový počet zpráv v bouři je součet geometrické řady

$$N = \sum_{n=0}^{\infty} R^n = \frac{1}{1-R} \quad \text{pro } R < 1.$$

Chování se láme v $R = 1$ — v *kritickém prahu*:

- $R < 1$ (*subkritický*): vlny se zmenšují, řada konverguje, bouře *s jistotou vyhasne*. Průměrně vznikne $1/(1-R)$ zpráv a je konec.
- $R = 1$ (*kritický*): každá vlna stejně velká, součet diverguje jen lineárně; systém balancuje na ostří nože.
- $R > 1$ (*superkritický*): vlny rostou geometricky, řada diverguje, bouře *exploduje* (a zastaví ji až vyčerpání seznamu nebo pád serverů).

Je to Galtonův–Watsonův větvící se proces, sepsaný v 19. století pro vymírání šlechtických příjmení. Táž rovnice, týž práh.



*Bouře odpovědí jako větvení se proces: jedna zpráva vyvolá v průměru R odpovědí, počet v n -té vlně = R^n .
 $R > 1$ exploduje (Bedlam DL3: ~15 mil. zpráv za ~hodinu), $R < 1$ vyhasne, $R = 1$ je práh mezi tím — tatáž matematika jako epidemie (R nula) a řetězová reakce (Galtonův–Watsonův proces: $R < 1$ útlum, $R > 1$ exploze).*

Hrdinský obraz: reply-all bouře jako větvení se proces. Svislá osa je logaritmická — proto se exponenciální exploze ($R > 1$) i exponenciální útlum ($R < 1$) vejdou do jednoho obrázku jako přímky opačného sklonu. Vodorovná čárkovaná čára $R = 1$ je kritický práh: hranice mezi bouří, co se sama krmí, a bouří, co se sama udusí.

Podívej se na obrázek, protože ten práh je vidět na první pohled. Když je R jen kousek nad jednou — řekněme každá zpráva vyvolá v průměru necelé dvě odpovědi —, křivka míří strmě vzhůru a po pár vlnách je z jedné zprávy pár set a za chvíli miliony. Když je R jen kousek pod jednou, křivka klesá a bouře se udusí dřív, než si jí kdokoli všimne. A přesně na jedničce je hrana: rozdíl mezi „nic se nestalo“ a „firma stojí“ nemusí být obrovský. Stačí, aby průměrná zpráva vyvolala o jednu odpověď víc nebo míň. Bedlam neměl R tisíc; nejspíš měl R jen mírně nad jednou — a to při desítkách tisíc lidí a mnoha vlnách úplně stačí na patnáct milionů zpráv.

To je, mimochodem, úplně tatáž matematika, jakou počítají epidemiologové. Jejich R_0 — základní reprodukční číslo — je totéž R : kolik lidí v průměru nakazí jeden nemocný. Pod jedničkou nákaza vyhasne, nad jedničkou se rozjede epidemie. A je to i matematika řetězové reakce v reaktoru a bombě: kolik nových štěpení vyvolá v průměru jedno štěpení. Pod jedničkou reakce utichá, nad jedničkou se lavinovitě rozbíhá. Tři na první pohled nesouvisející věci — e-maily, nemoci, atomová jádra — poslouchají jeden a tentýž zákon větvení. To je ta krása, kvůli které matematika stojí za námahu: jednou pochopíš práh $R = 1$ a rozumíš najednou třem světům.



A teď ta část, kvůli které je věž v knize o *inženýrství*, ne v učebnici pravděpodobnosti: R není osud. R je *součin* věcí, které se dají navrhnout. Zhruba

$$R \approx p \times m,$$

kde m je počet příjemců, kteří zprávu uvidí, a p je pravděpodobnost, že jeden z nich odpoví. Chceš $R < 1$? Sniž kterýkoli činitel:

- *limit příjemců* (m dolů): strop na velikost seznamu — přesně to Microsoft udělal jako první opatření po Bedlamu.
- *moderace* (p dolů): odpověď musí někdo pustit dál, ne se rozešle sama.
- *tlumící zpoždění*: rate limit — bouře nemá čas nabrat vlny, než ji někdo zastaví (→ kap. 21).

Design nemění to, že platí $N_n = R^n$. Mění *hodnotu* R — a tím rozhoduje, na které straně prahu skončíš.

Tohle je nejužitečnější věta celé věže, tak ji řekněme natvrdo: lavinu neporazíš tím, že budeš doufat, že se lidé budou chovat rozumně. Porazíš ji tím, že *snížíš R pod jedničku* — návrhem, ne přáním. Ubereš příjemce, přidáš moderaci, vložíš zpoždění. Každá z těch pák tlačí R dolů a existuje bod, za kterým se žádná bouře fyzicky nemůže rozjet, ať se lidé chovají jakkoli. Není to o disciplíně účastníků. Je to o čísle, které jsi jim nastavil dřív, než začali psát.

Zpátky ke stroji

Vrať tu rovnici na model, protože přesně tam začne pracovat naostro. Když necháš AI něco *rozeslat* — dvaceti lidem, do všech větví repozitáře, na každý řádek tabulky —, ptej se úplně stejně jako u e-mailu: *kolik akcí vyvolá jedna akce?* Když necháš agenta „projít všechny soubory a opravit je“, a každá oprava spustí přebuildování, které spustí testy, které pošlou zprávu, která spustí dalšího agenta — právě jsi postavil větvící se proces a máš svoje R . Jestli je nad jedničkou, nepostavil jsi automatizaci. Postavil jsi lavinu, jen zatím nevybuchla, protože je malý seznam.

A tady se ukazuje, proč je zesilovač bez směru nebezpečnější než kdy dřív. E-mail rozešle tvůj zmatek dvaceti lidem, kteří ho aspoň přečtou a někdo z nich zabrzdí. Model a agent rozešlou tvůj zmatek *tisícům operací* za vteřinu, bez lidského oka mezi nimi, s R , které jsi nespočítal, protože tě to nenapadlo počítat. Kritický práh se nezměnil — pořád je na jedničce. Změnila se rychlost, s jakou ho překročíš, a počet vln, které proběhnou, než si toho všimneš.

→ kap. 21: rate limity, alerty a monitoring jsou přesně protilavinové zábrany — nastavené číslo, které drží R pod prahem, když spíš.

→ kap. 3: až lavina proběhne, budeme v tom zpětně hledat vzory a příběh — i tam, kde byla jen slepá geometrie větvení.

Proto tahle kapitola končí u čísla, ne u dobré rady. „Buď opatrný“ je R , které nikdo nezměřil. Zato „než to spustím, spočítám, kolik akcí každá akce vyvolá, a jestli je to víc než jedna, přidám brzdu“ — to je snížení R pod práh, návrhem, předem. To je celý rozdíl mezi inženýrem a člověkem, který napsal do Bedlamu DL3, že chce být vyřazen ze seznamu.

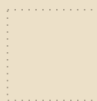
Co si ze sedmé kapitoly odnést

Poskládej to dohromady, protože obraz je jednoduchý a jedno číslo ho drží. Nástroj je zesilovač a zesilovač nezná směr: dobrý proces zesílí na dobrý výstup, chaos na větší chaos, a nezvolí, co z toho nastane — to určuje vstup. E-mail to dělá padesát let a pořád na to naletíme, my všichni; model to dělá stejně, jen s mnohem vyšším ziskem, takže z nejasného zadání dostaneš delší, zdvořilejší a početnější verzi téhož nejasna. Obrana není ztlumit zesilovač, ale dát na vstup protokol: rozdělit výstup na tvrzení (ověř), akce (pojistit) a ozdobu (zahodit). A pod tím vším leží jedna rovnice: $N_n = R^n$, kde R je počet odpovědí či akcí vyvolaných jednou. Nad jedničkou exploze, pod jedničkou útlum, na jedničce ostří nože — a R se dá návrhem snížit (míň příjemců, moderace, zpoždění). Tatáž matematika jako epidemie a řetězová reakce. Kdo tohle ví, přestane doufat v rozumné chování a začne nastavovat číslo.

A žádný posměch nad těmi, co e-mail neumí — protože ho neumíme my všichni, já první. Bedlam nezhroutili hlupáci; zhroutili ho chytrí lidé s dobrým úmyslem a zesilovačem bez brzdy. Přesně proto stavíme brzdy: ne protože jsme neschopní, ale protože zesilovač je mocný a mocný nástroj bez protokolu zesílí i tu nejlidštější malou chybu do bouře, kterou nikdo nechťel.

■ DEMO

`lesk-a-bida-vibecodingu.gaussalgo.com/d/reply-all-storm`



Bouře odpovědí naživo: táhni posuvníkem R (kolik odpovědí vyvolá jedna zpráva) a sleduj, jak se lavina v grafu organizace buď udusí (R pod jedna), drží ($R = 1$), nebo exploduje (R nad jedna). Vedle: zesilovač — pusť týž vstup při zisku $1 \times$ a $1000 \times$ a podívej se, jak kvalitní proces dá čistý výstup a chaotický dá lavinu, aniž by se stroj cokoli změnil.

„Jak bych poznal, že se pletu?“

Konkrétní test k sedmé kapitole, a je to jedno číslo: *než pošleš nebo vygeneruješ hromadnou věc — mail do velkého seznamu, agenta přes všechny soubory, skript přes celou databázi — zeptej se, kolik odpovědí nebo akcí každá jedna vyvolá*. Když víc než jednu, staviš lavinu, i když to zatím nevybuchlo. Sniž R pod jedničku dřív, než zmáčkneš enter: uber příjemce, přidej potvrzení, vlož brzdu. A jestli chceš tuhle kapitolu shodit: tvrdím, že u příští hromadné akce, kterou pustíš bez tohoto odhadu, tě aspoň jednou za rok překvapí kaskáda, kterou jsi nečekal — protože R vyšlo nad jedna a tys ho nespočítal. Když se ti to nikdy nestane, buď máš R pod prahem už ze zvyku, nebo jsi nikdy nic velkého nespustil. To první je přesně cíl.