Cause and effect

Every action has an outcome, big or small.This outcome may be predictable:

Deterministic: When the outcome is unique (we know in advance what will happen), the action is called **deterministic**, meaning that the outcome is predetermined by the action.

Random: When several outcomes are possible, the result of the action is–to some extent–random. The term **stochastic** is applied when the randomness of the action has some probabilistic nature.

Unknown: When it is unclear what the outcome of an action will be, it does not make it deterministic (not as obvious as it seems) nor stochastic; we merely do not know. Numerous philosophers tried to deal with the cause–effect relationship; we will leave it to them.

Deterministic vs. Stochastic

The basic distinction is: **one** or **many**. Reality makes the distinction fuzzy:

- I am about to unlock a door and I have the right key (guaranteed). Looks deterministic: the outcome of the **action of unlocking** is an **unlocked door**. True until you experience a lock jam.
- The action is: I am about to toss a coin. Is it deterministic or stochastic? It depends on the outcome of interest:
  - Will it fall on the ground?
  - Where will it fall?
  - Will it fall tails up?
  - Will it disappear into thin air while falling?

Process

A **dynamic system** is any closed **whole** that changes in time. See wikipedia if this definition intrigues you.

A **process** is every sequence of changes of the state of a system. The **changes** forming a process are called **events**.

Grouping of events into processes is an arbitrary operation which depends on the goals of the observer.

Let us look at a simple example: a single–window hamburger stand.

From the point of view of customer flow, the system looks approximately like this:

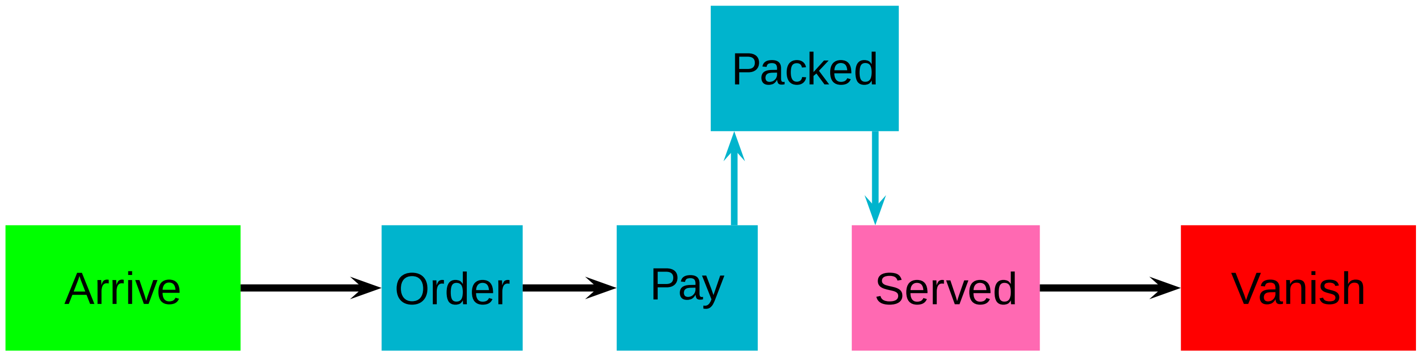
- The **input process** generating new customers (more precisely: **customer arrivals**). It is convenient to associate this process with a **queue**.
- The **server process** which disposes of customers, presumably after serving them.

Make sure to understand in full why the **actual work** done by the system (serving customers) is not included in any process and seems to be absent from this system[a]

---

[a] It is not really absent; more correctly: ignored.

The simplistic view of the system hides a few potentially relevant details.



How many processes do we have here? It helps to introduce the notion of the **state** of a customer. An event that changes the state of a customer must be part of a process.

For the system above, the states are:

$$Arrive \longrightarrow Order \longrightarrow Pay \longrightarrow Vanish$$

The moving parts

Some of the processes can be viewed as deterministic, i.e. the events forming them occur at predetermined[a] intervals. Additionally the output process is by definition deterministic: a customer that reaches the output process simply disappears from the system (and becomes“just a statistic”).

Conversely, some processes are made of events that do not occur at predetermined intervals. In any decent restaurant, the time elapsing between

- the moment the food is served to a customer (changing the customer’s state from **WaitingForFood** to **EatingFood**)
- and the moment the customer chooses to enter the next state (be it **WantingToPay** or **RunWithoutPaying** or **ExitOffended** etc.) is highly unpredictable. Note that this moment is not even related to the action of eating.

Processes influenced by randomness are called **stochastic**.

[a] Not necessarily fixed.

Customer arrivals

The input process is not always stochastic, but it usually is. Certainly, arrivals to a shoe polish stand are unpredictable. The same is true of disk i/o requests submitted to a disk controller.

When customers arrive at random moments, the arrivals form (by definition) a random variable that follows some probability distribution. Later we will look at some options; now is the time to emphasise that it creates counter–intuitive stochastic effects.

Exhibit 1

In a small town there is a single dentist extracting teeth. This is an old–fashioned fellow who takes exactly 60 minutes per customer (if the extraction is too simple, the dentist fills the time with jokes, etc.).

Customers arrive at intervals that are uniformly distributed with a minimum of 0 minutes and a maximum of 120 minutes (so, with an average of 60 minutes).

We have a simple single–server system with a probabilistic interarrival time of 60 minutes **on average** and a deterministic service time of 60 minutes.

What are the statistics describing the waiting time for service?

To simplify matters, we decide to compute the maximum length of the waiting queue in the dentist's office (which gives a bound on the waiting time).

Not knowing any better, we write a short program that simulates the work of the office for 1000 minutes. The maximum queue length is 2. Try 5,000: the maximum is 8. Try 10,000: the maximum still is 8.

10,000	8
15,000	10
20,000	19
50,000	19
100,000	20
200,000	34
300,000	50
400,000	72
500,000	90
700,000	119
1,000,000	119
5,000,000	119
10,000,000	257

10,000	8	11	14
15,000	10	16	14
20,000	19	16	14
50,000	19	16	21
100,000	20	16	21
200,000	34	34	24
300,000	50	44	24
400,000	72	53	25
500,000	90	53	61
700,000	119	53	122
1,000,000	119	128	131
5,000,000	119	230	421
10,000,000	257	230	421

Stochastic effects

Many authors use the term “stochastic effect” to denote the presence of randomness in a process.

However, it is not the presence of randomness alone that is worth attention: it rather is the fact that this randomness can lead to very unexpected outcomes in individual trials.

As an example, take the **random walk** (also known as **drunkard’s walk**), a simple algorithm in which a point moves along an infinite straight line, every time unit choosing at random to move forward or backward one step.

The drunkard starts at location 0 and chooses at random to move to $+1$ or -1. The toss of a coin simulates well enough the randomness of choice.

Everyone will agree that after any number of steps n the expected location of the drunkard is 0.

Reality show

Any simulation shows that it **almost never** is the case.

A mathematical derivation shows that the expected position of the drunkard after **n** steps is $\sqrt{\frac{2\times n}{\pi}}$ away from 0.

Stochastic effects

Wherever random numbers are used, one should expect stochastic effects although they seldom show in a way as conspicuous as in the drunkard's walk.

There are two basic kinds of stochastic effects:

- If a **pd** has a mean μ then the value of any random variate $\mathcal{X}$ drawn this **pd** is not likely to be μ. All that can be claimed is that if $\mathcal{X}_1, \ldots, \mathcal{X}_n$ are drawn from this **pd**, then
$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathcal{X}_i = \mu$$
The same applies to all other statistics.

- If a **pd** has a range of possible values $<a, b>$, than a random variate drawn from this **pd** can take any value in the range, not just values close to the mean. A situation when a particularly unlikely (but not impossible) outcome disturbs greatly the life of a model is called the "appearance of a **black swan**" and is bound to disturb simulation results.

Black swans

A common example of black swans comes from the exponential distribution (essentially: customer inter–arrival times). Regardless of the mean inter–arrival time, the maximum possible value is ∞, hence a simulation of a hamburger stand with a mean inter–arrival time of 2 minutes might yield a case when the next customer **will never arrive** no matter how long the simulation runs.

To avoid being misled by black swans, a **large number** of independent experiments must be conducted.