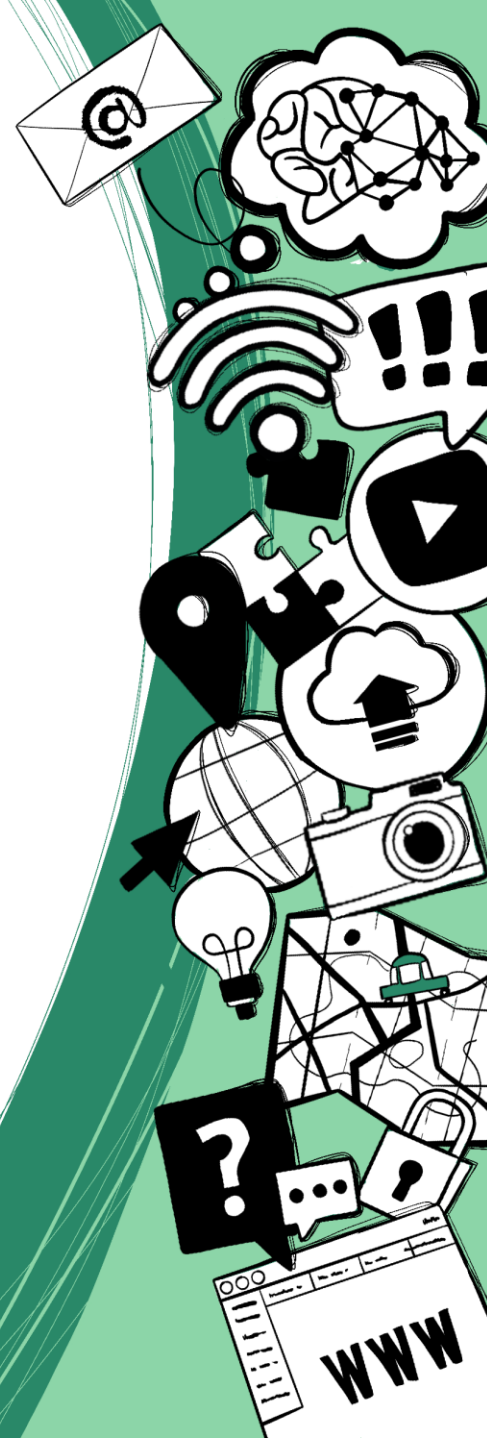
ENARIS

Labyrinth



mobilis
interaktív kiállítási központ

TU Graz

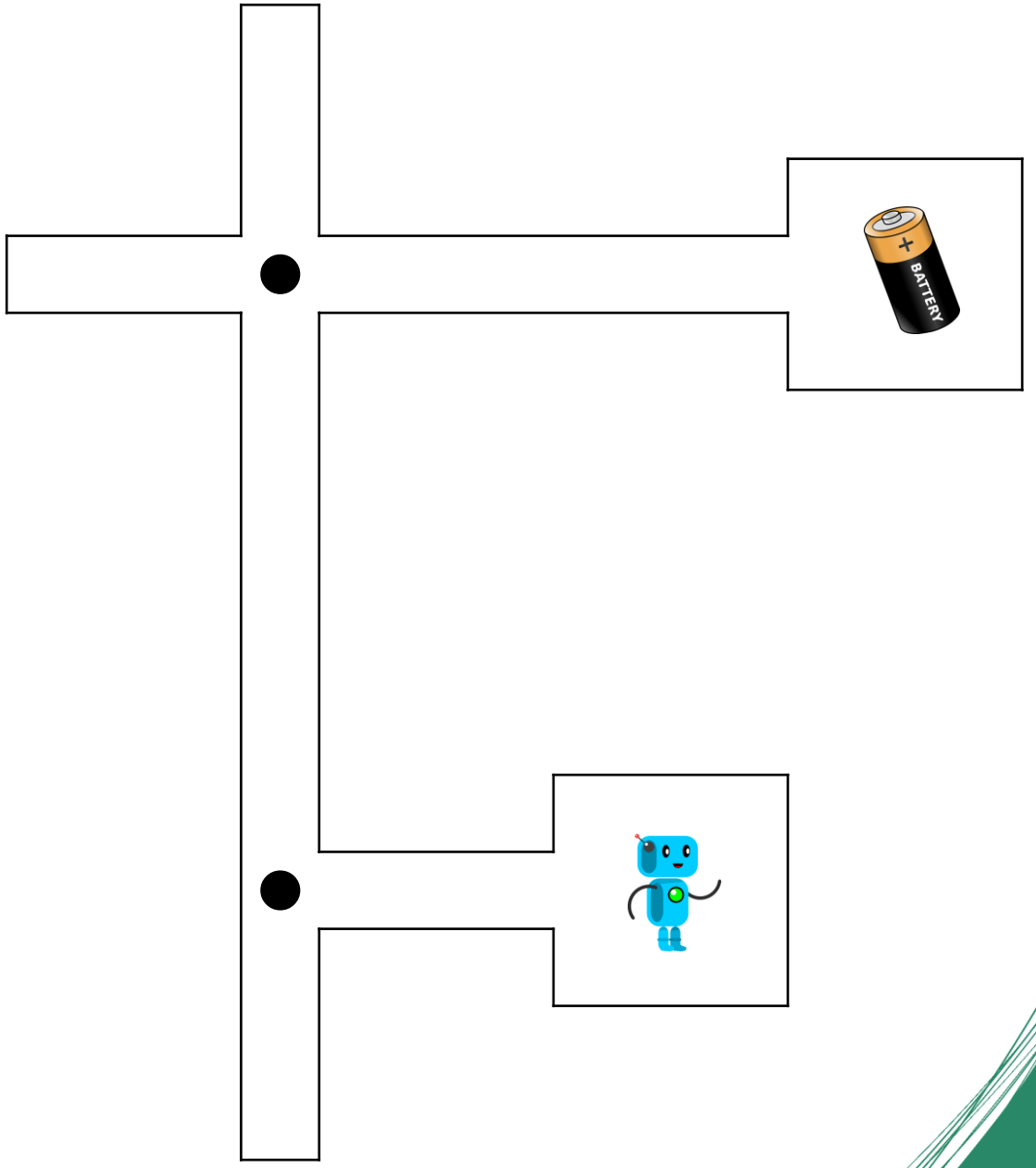
OESTERREICHISCHE COMPUTER GESELLSCHAFT ®
AUSTRIAN COMPUTER SOCIETY

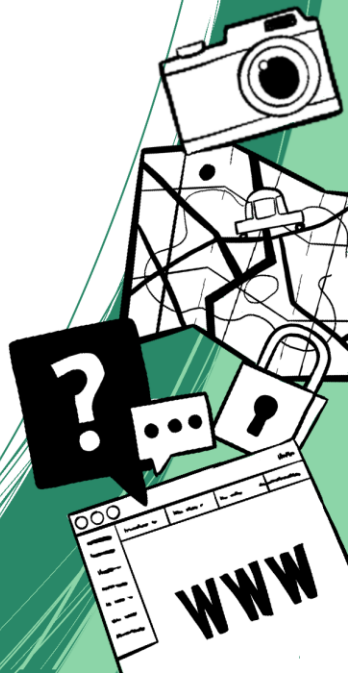
Interreg
Austria-Hungary 2014-2020
European Union – European Regional Development Fund

Grundregeln

Das Spiel

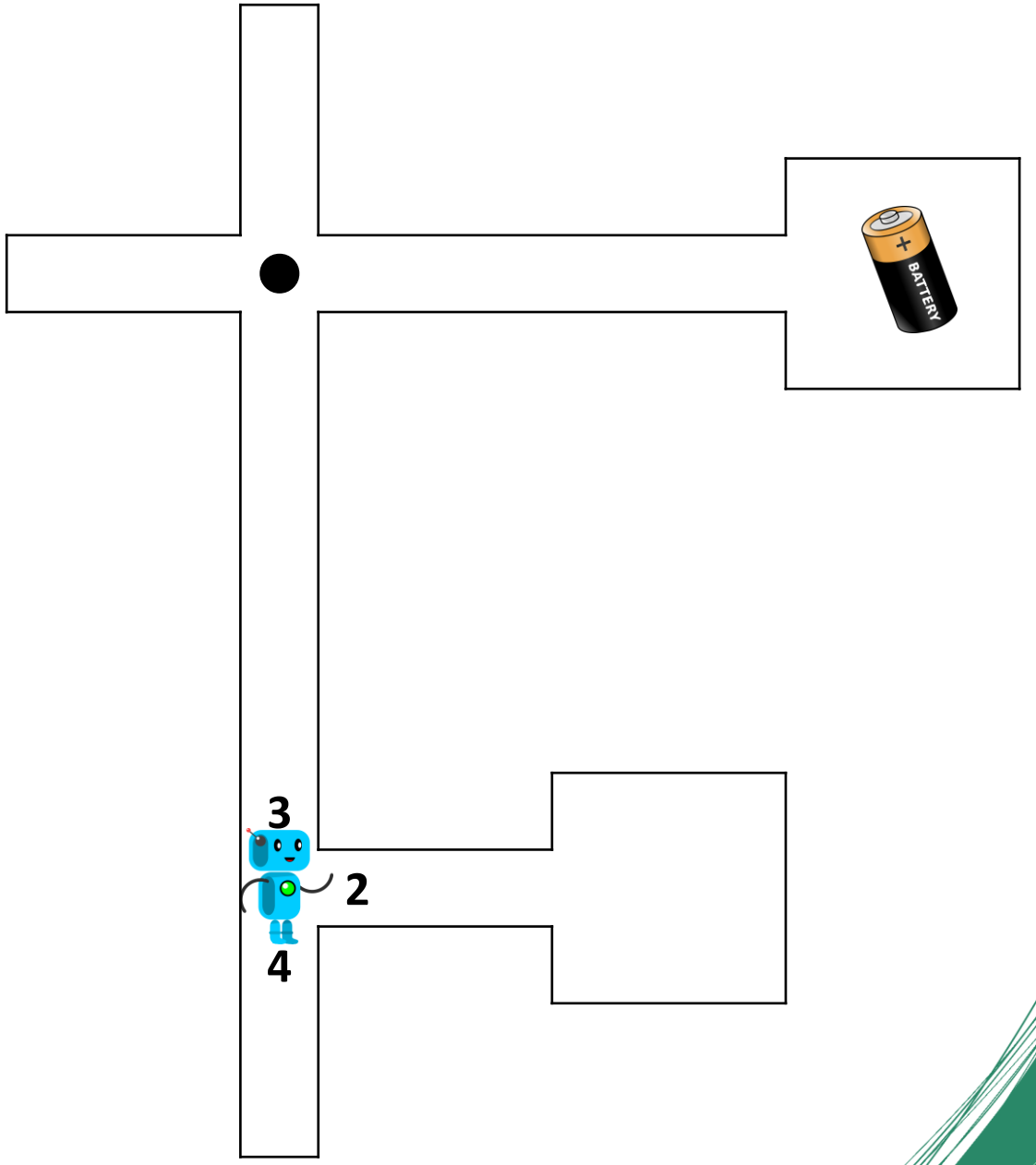
- **Agent**: Roboter
- **Aktionen**: nimm einen Weg
- **State**: Labyrinth, Positionen, Q-Werte
- **Belohnungen**: Zahlen entlang der Wege
- **Ziel**: einen Weg zur Batterie finden

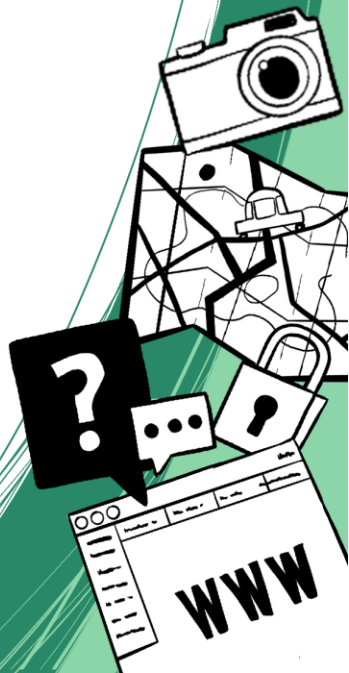




Das Spiel

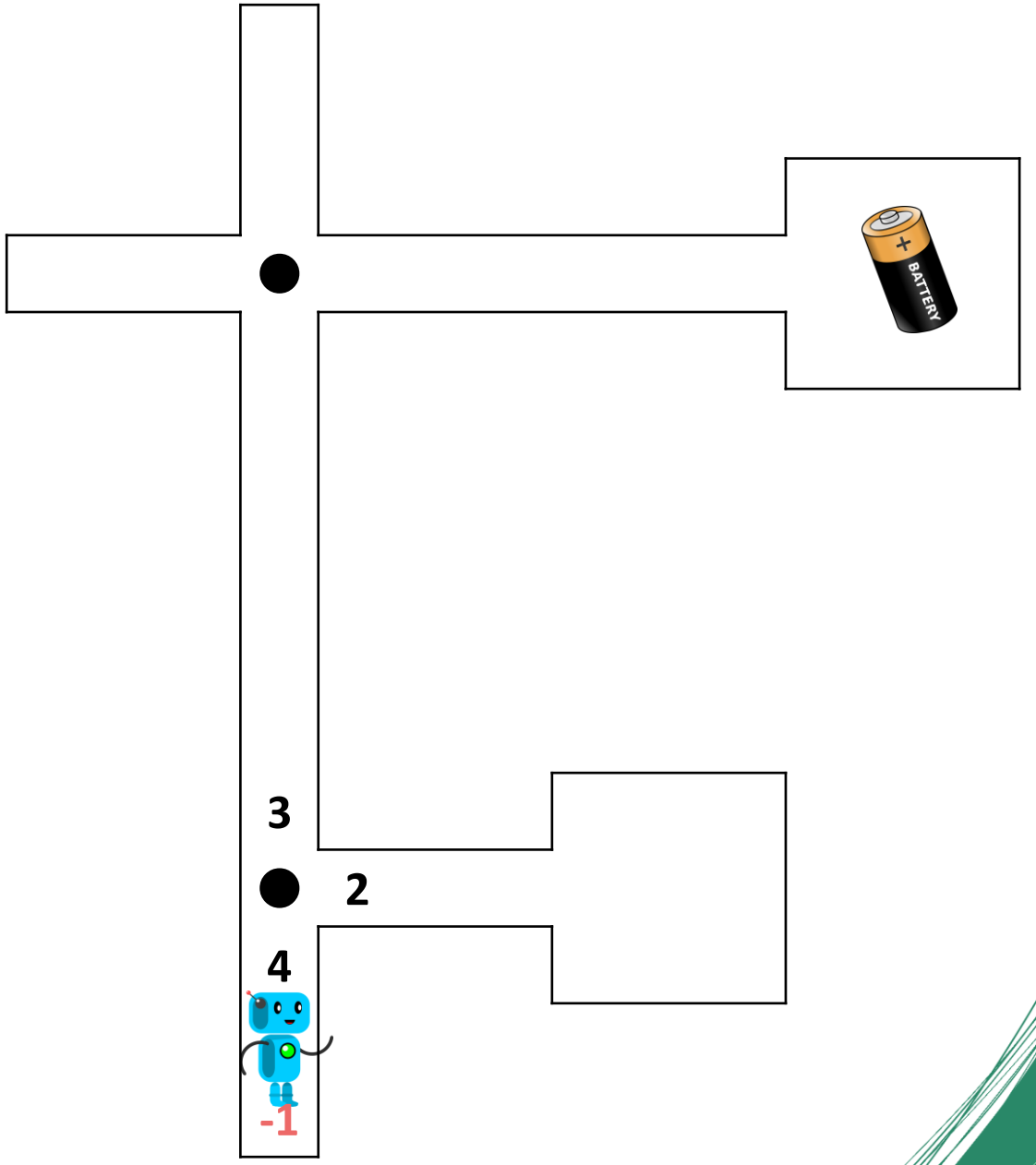
- Immer, wenn der Roboter auf eine **neue Kreuzung** stößt, schreibt er eine **zufällige Zahl** (**würfeln**) auf jeden **möglichen Weg**
- Der Roboter geht dann den Weg mit der **höchsten Zahl**

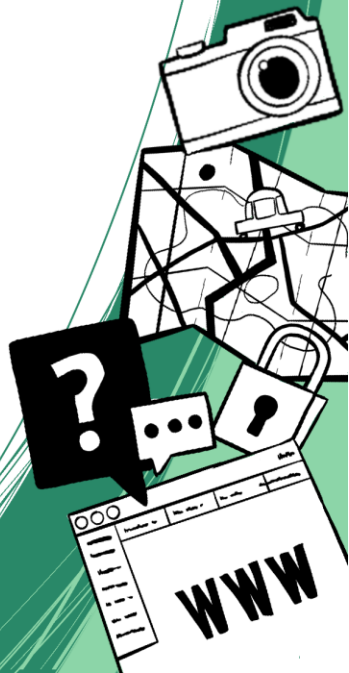




Das Spiel

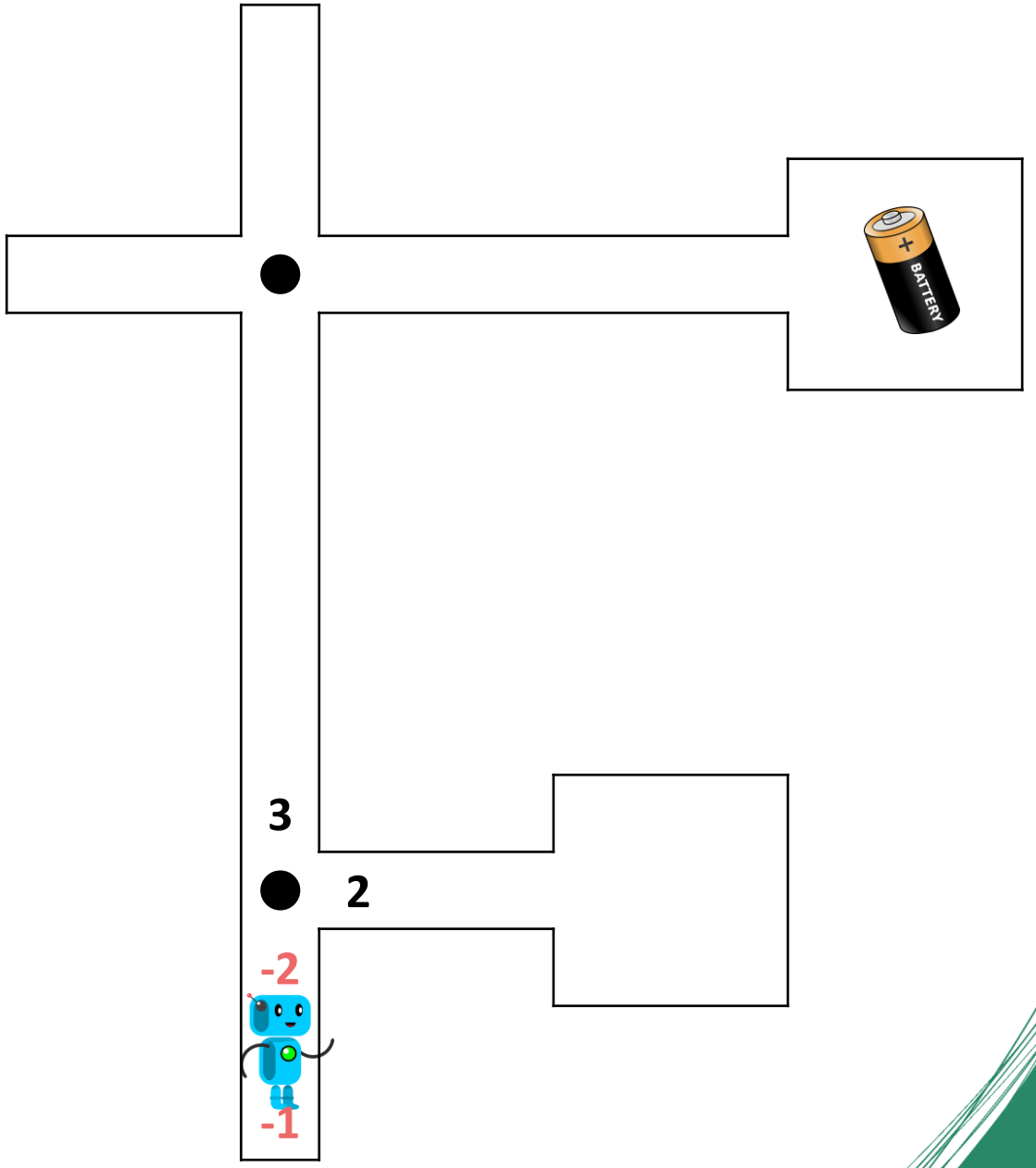
- Wenn der Roboter in eine **Sackgasse** kommt, schreibt er **-1**

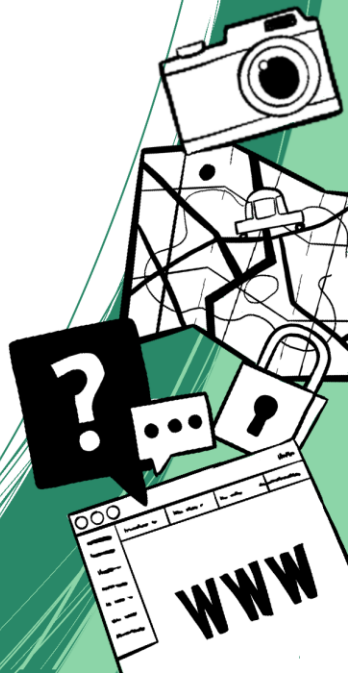




Das Spiel

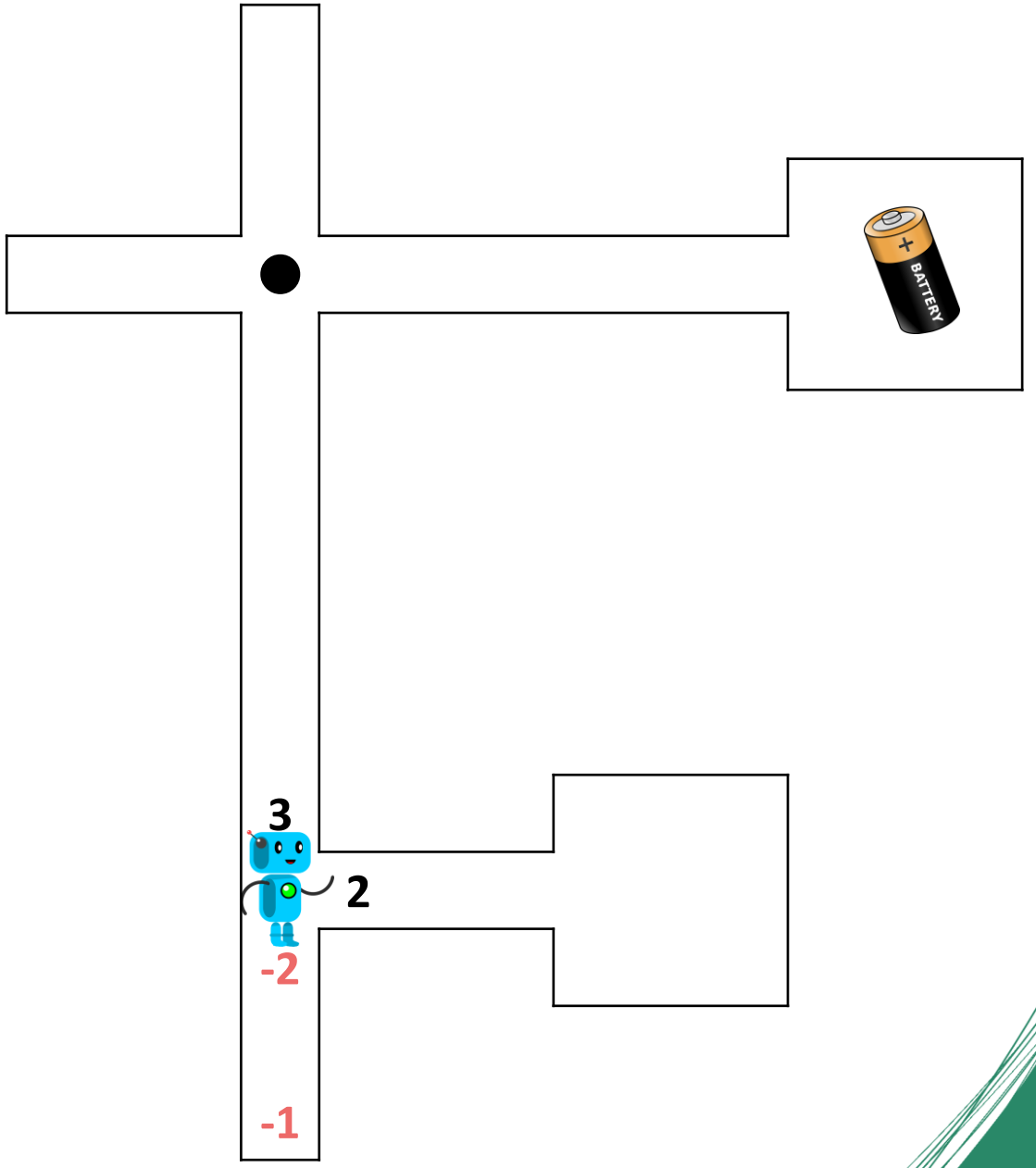
- Wenn der Roboter in eine **Sackgasse** kommt, schreibt er **-1**
- Dann wird die **Zahl** des Weges, **von welchem der Roboter kam**, auf den neuen **höchsten wert** (-1), **minus eins** (-1-1=-2), gesetzt

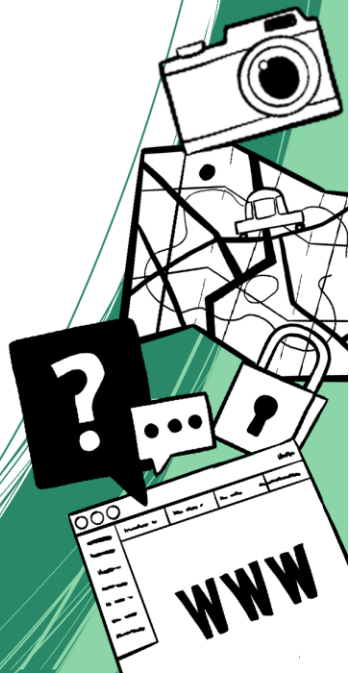




Das Spiel

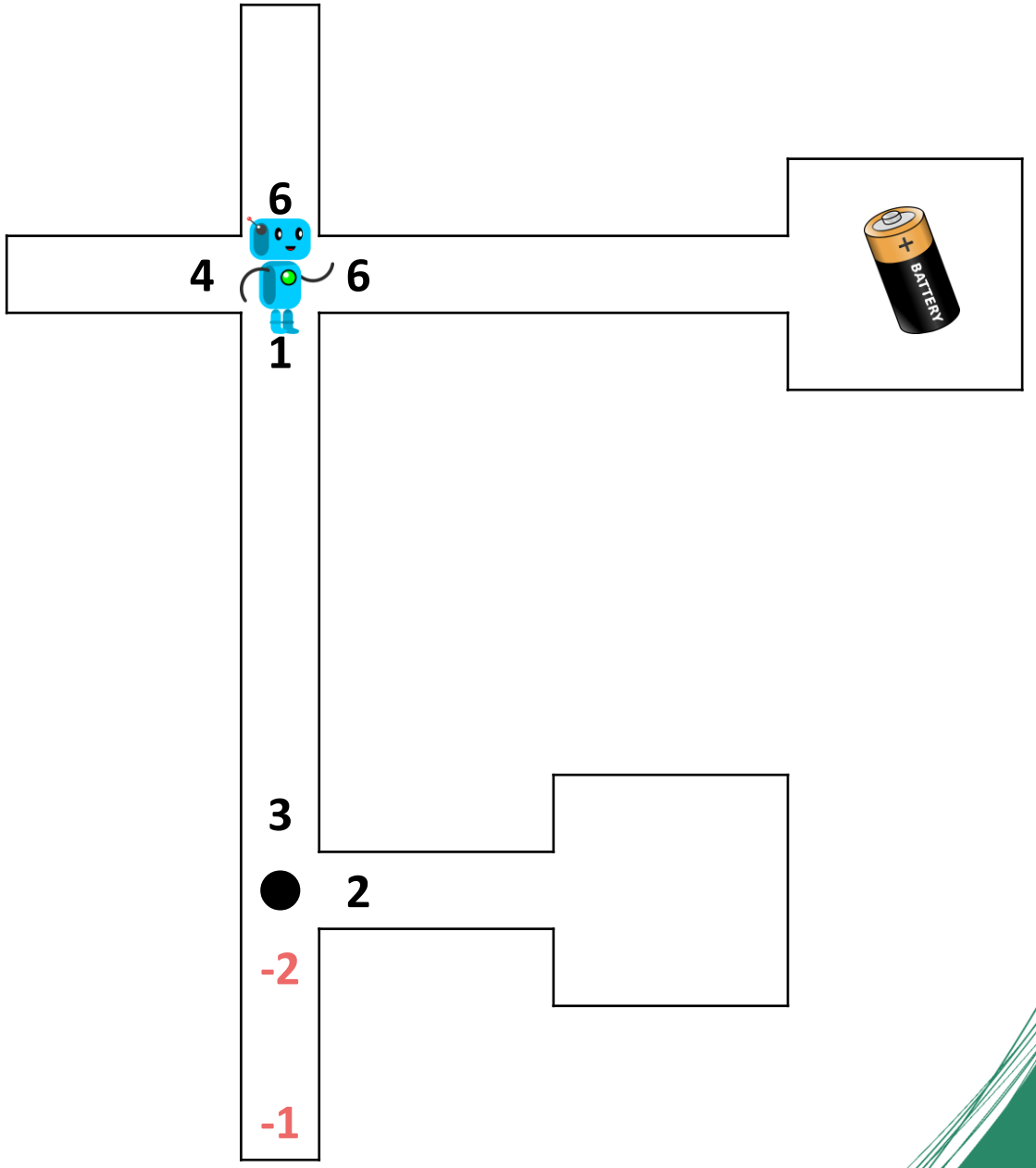
- Wenn der Roboter in einer **Sackgasse** war, geht er wieder zur **letzten Kreuzung** zurück und nimmt den neuen Weg mit der größten Zahl





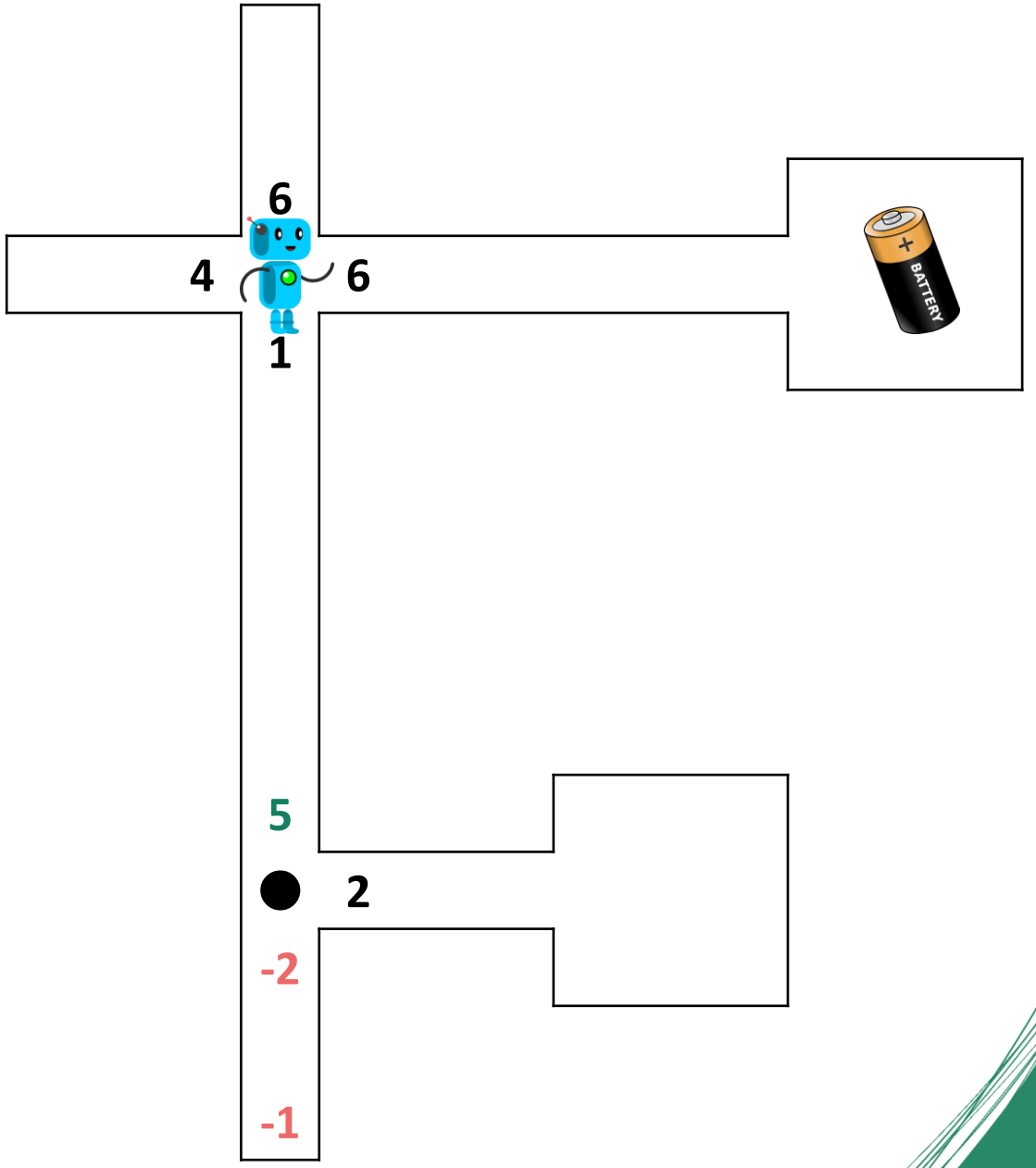
Das Spiel

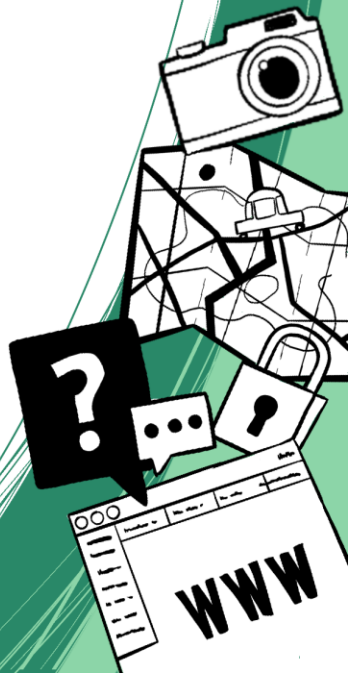
- **Neue Kreuzung** -> zufällige Zahlen



Das Spiel

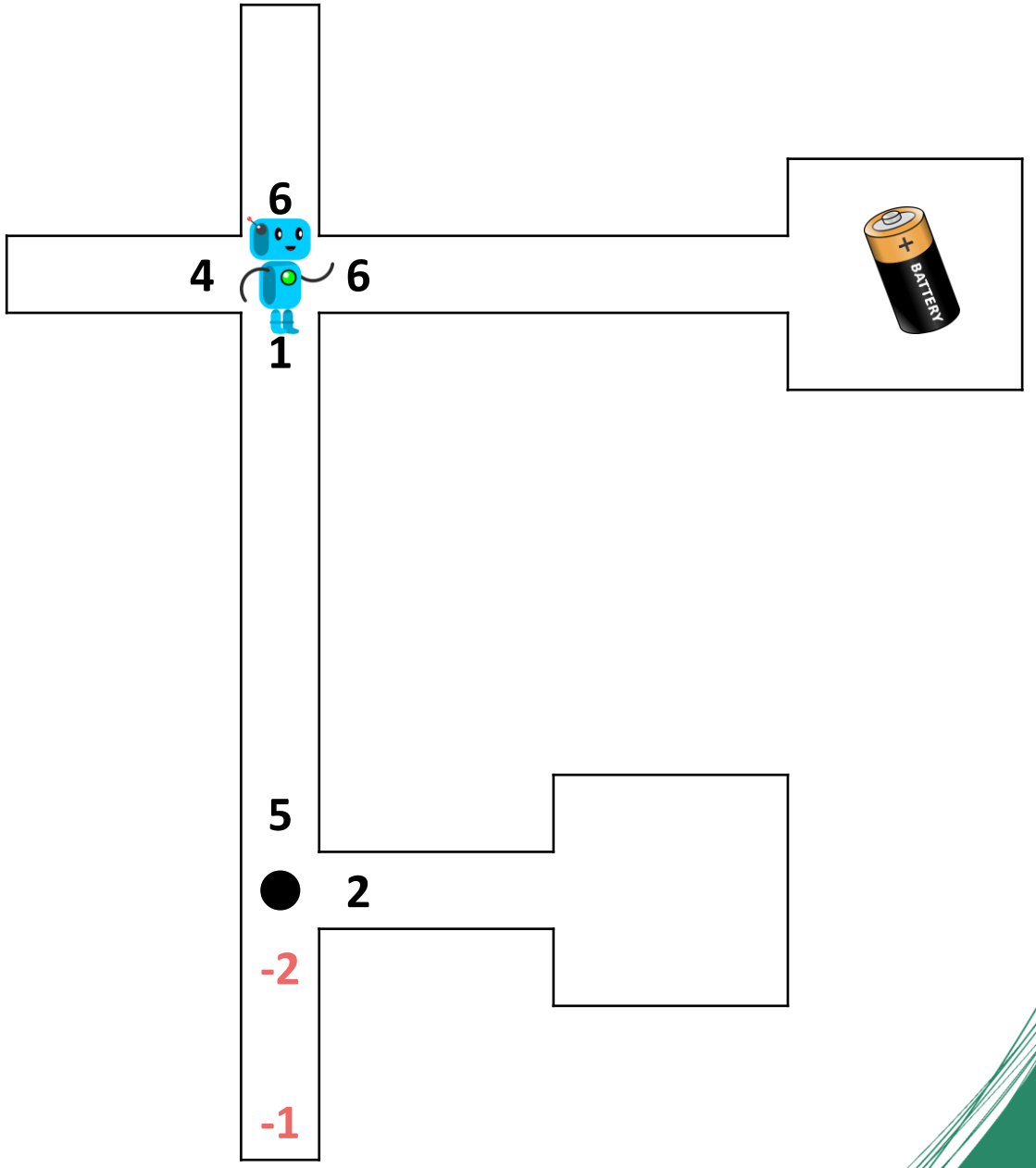
- **Neue Kreuzung** -> zufällige Zahlen
- **Update** die Zahl des Weges auf die neue größte Zahl (6) minus eins(6-1=5)

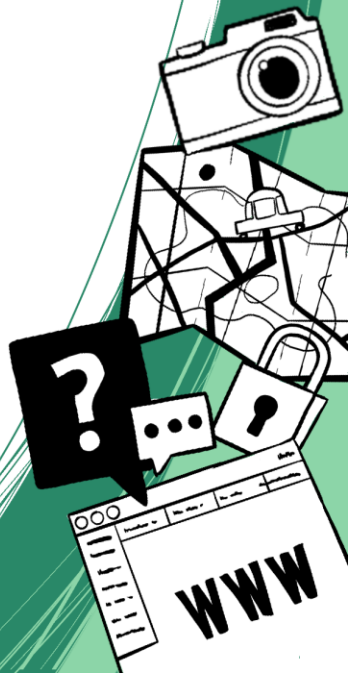




Das Spiel

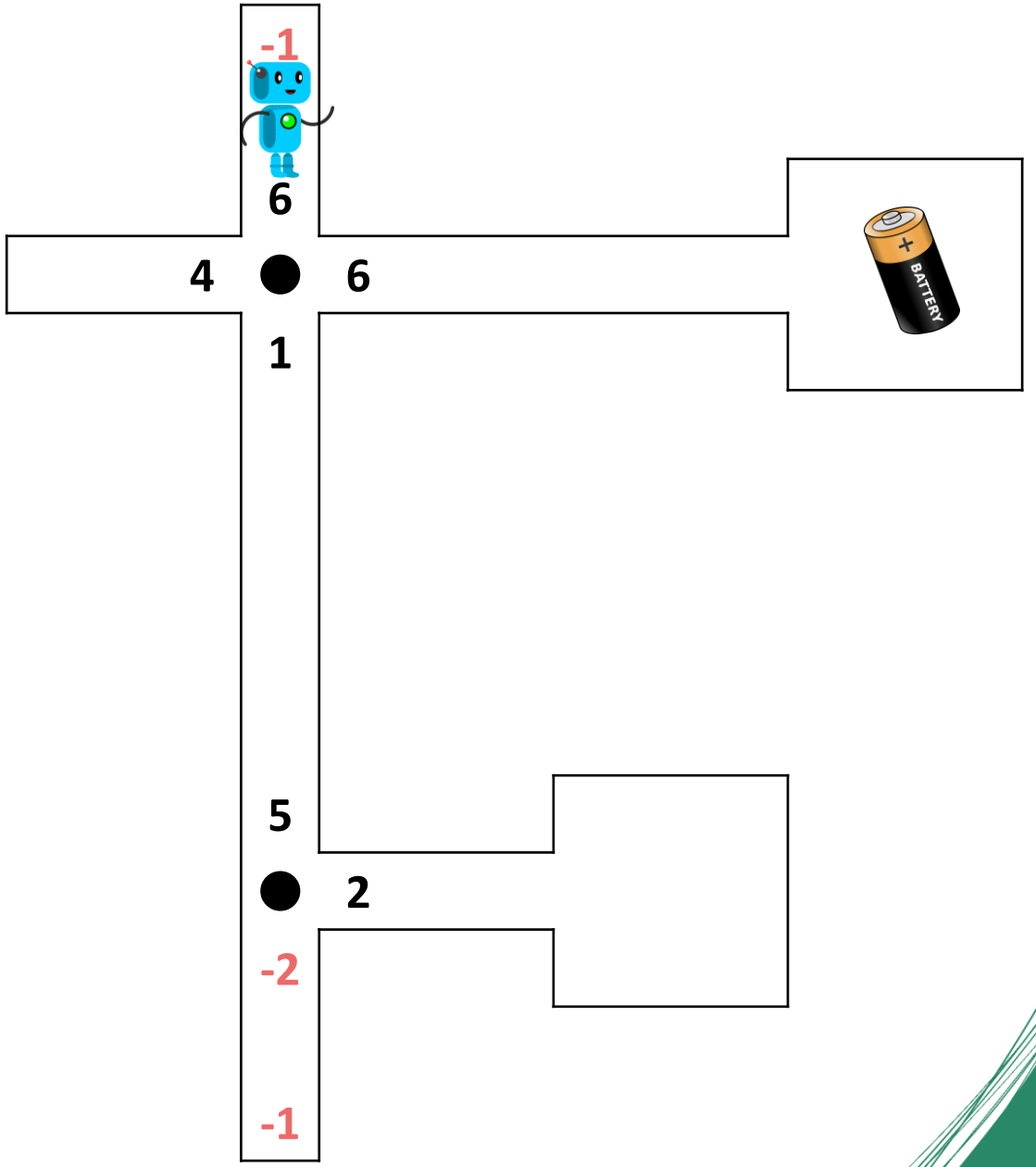
- **Neue Kreuzung** -> zufällige Zahlen
- **Update** die Zahl des Weges auf die neue größte Zahl (6) minus eins(6-1=5)
- Bei **zwei oder mehr** gleich großen Zahlen nimmt der Roboter **zufällig** einen der Wege (z.B. hinauf)

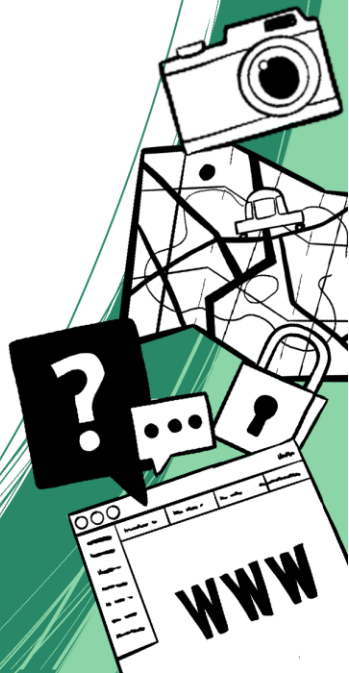




Das Spiel

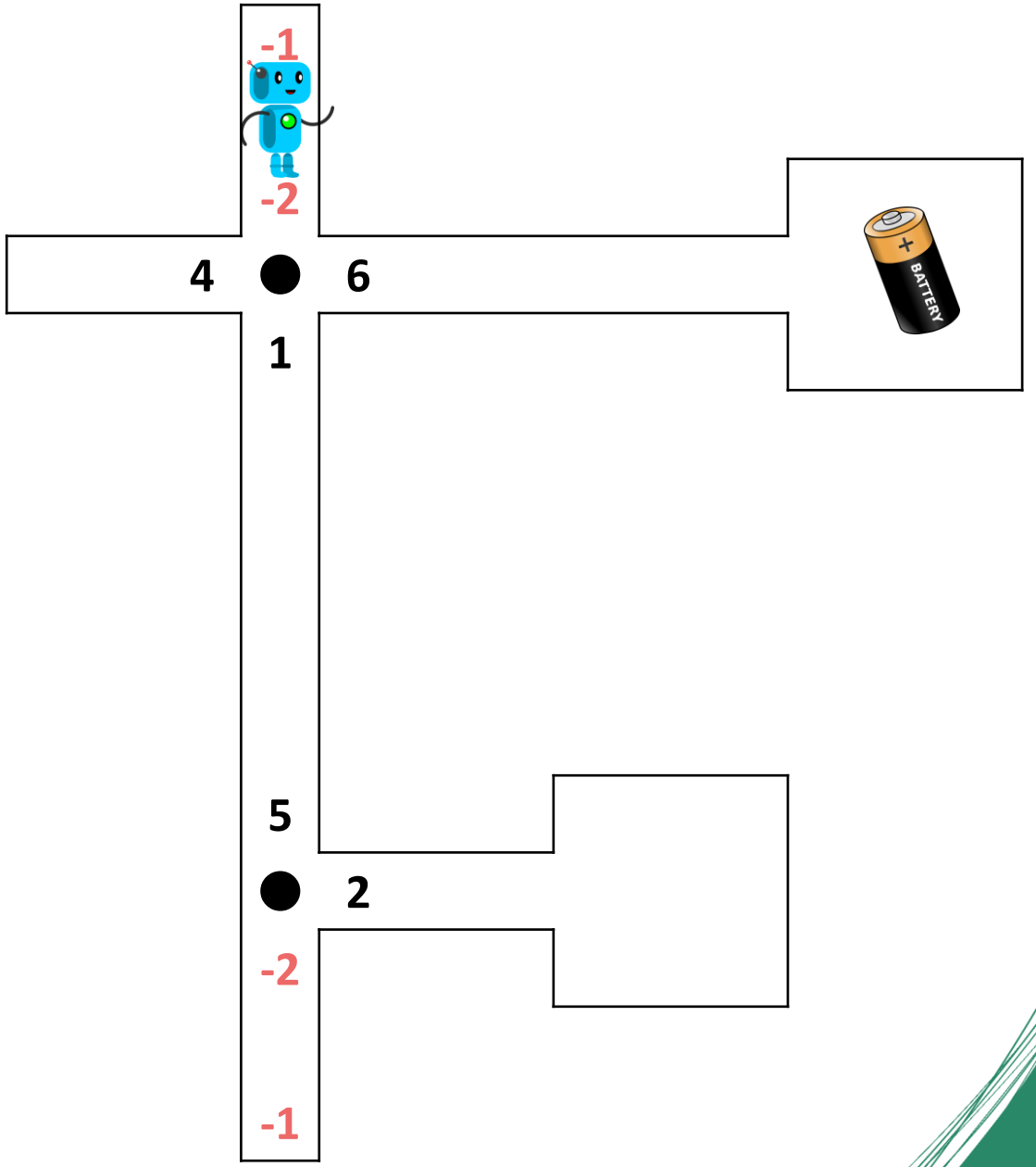
- Der Roboter **macht so weiter**, bis er sein **Ziel** erreicht hat

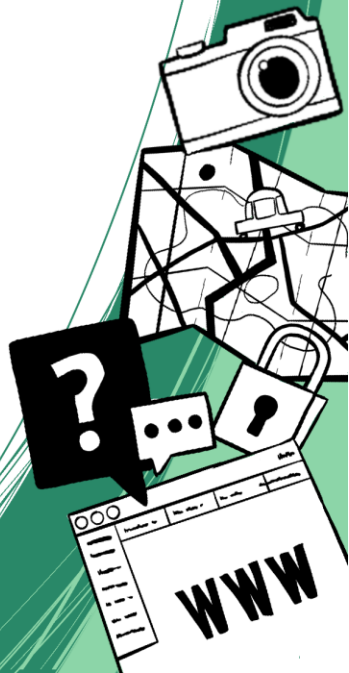




Das Spiel

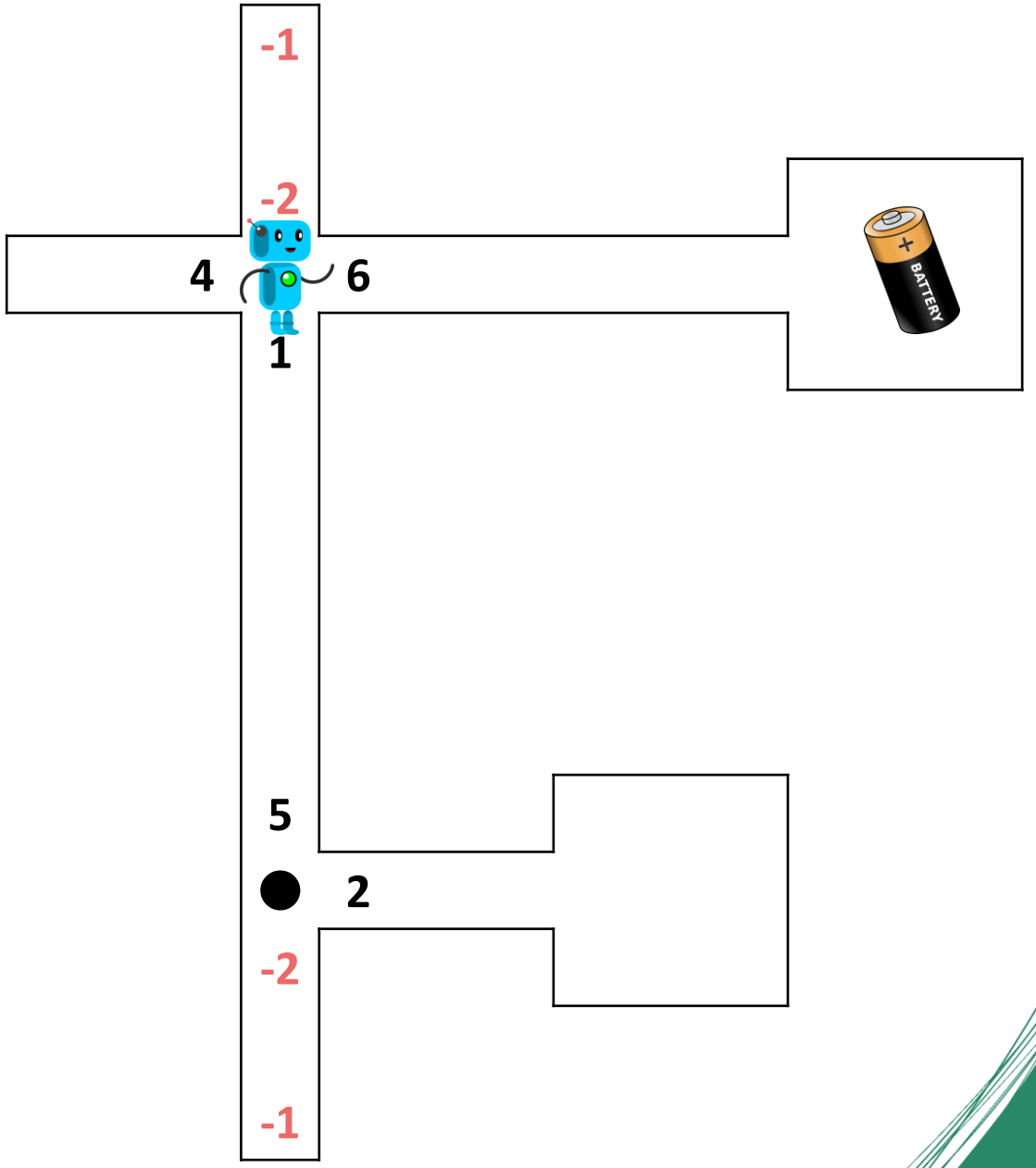
- Der Roboter **macht so weiter**, bis er sein **Ziel** erreicht hat

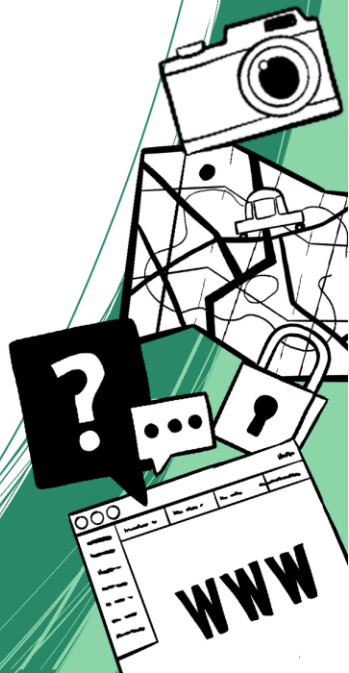




Das Spiel

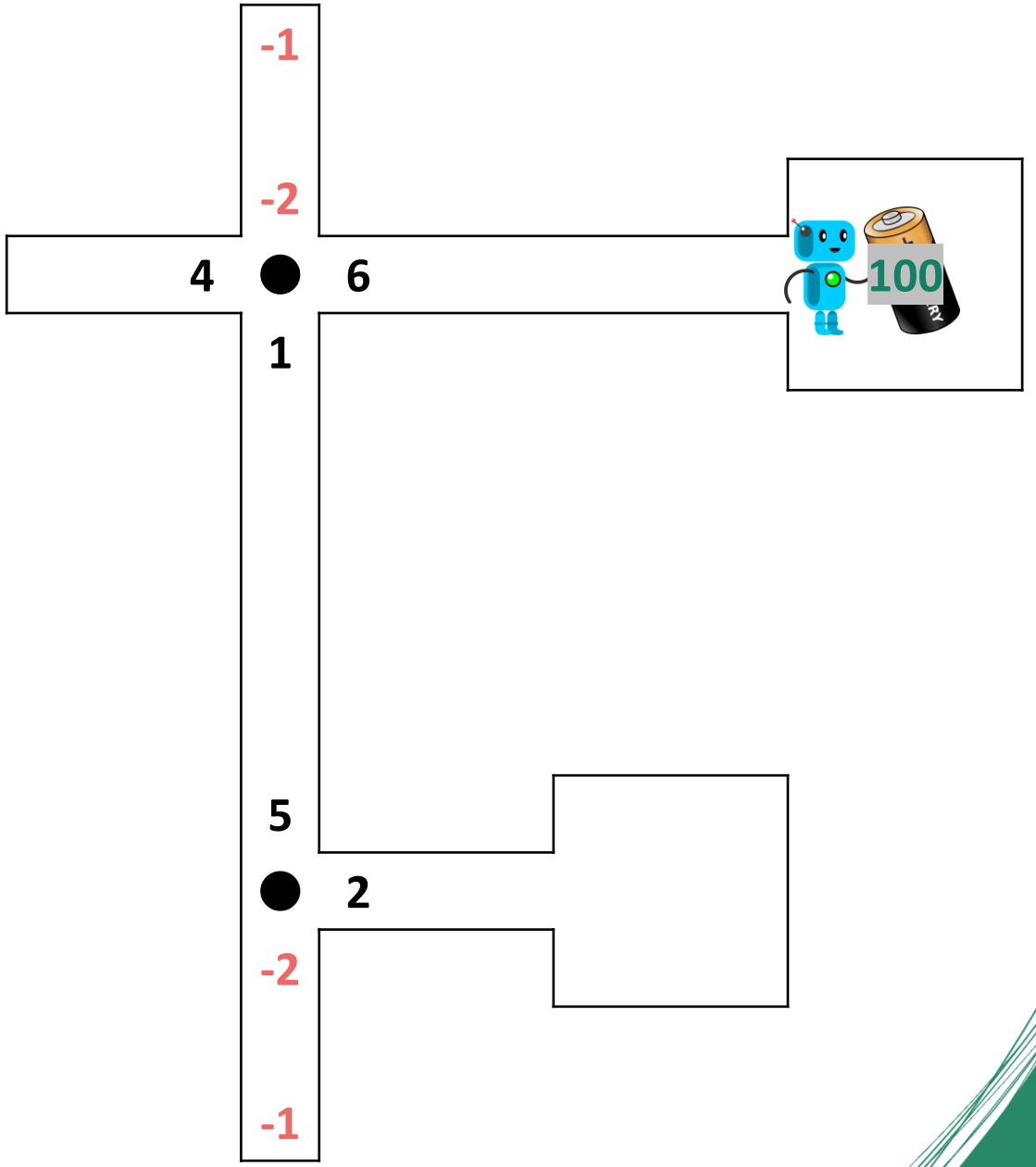
- Der Roboter **macht so weiter**, bis er sein **Ziel** erreicht hat

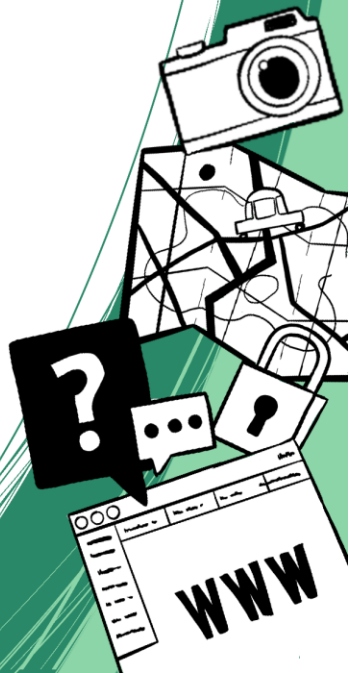




Das Spiel

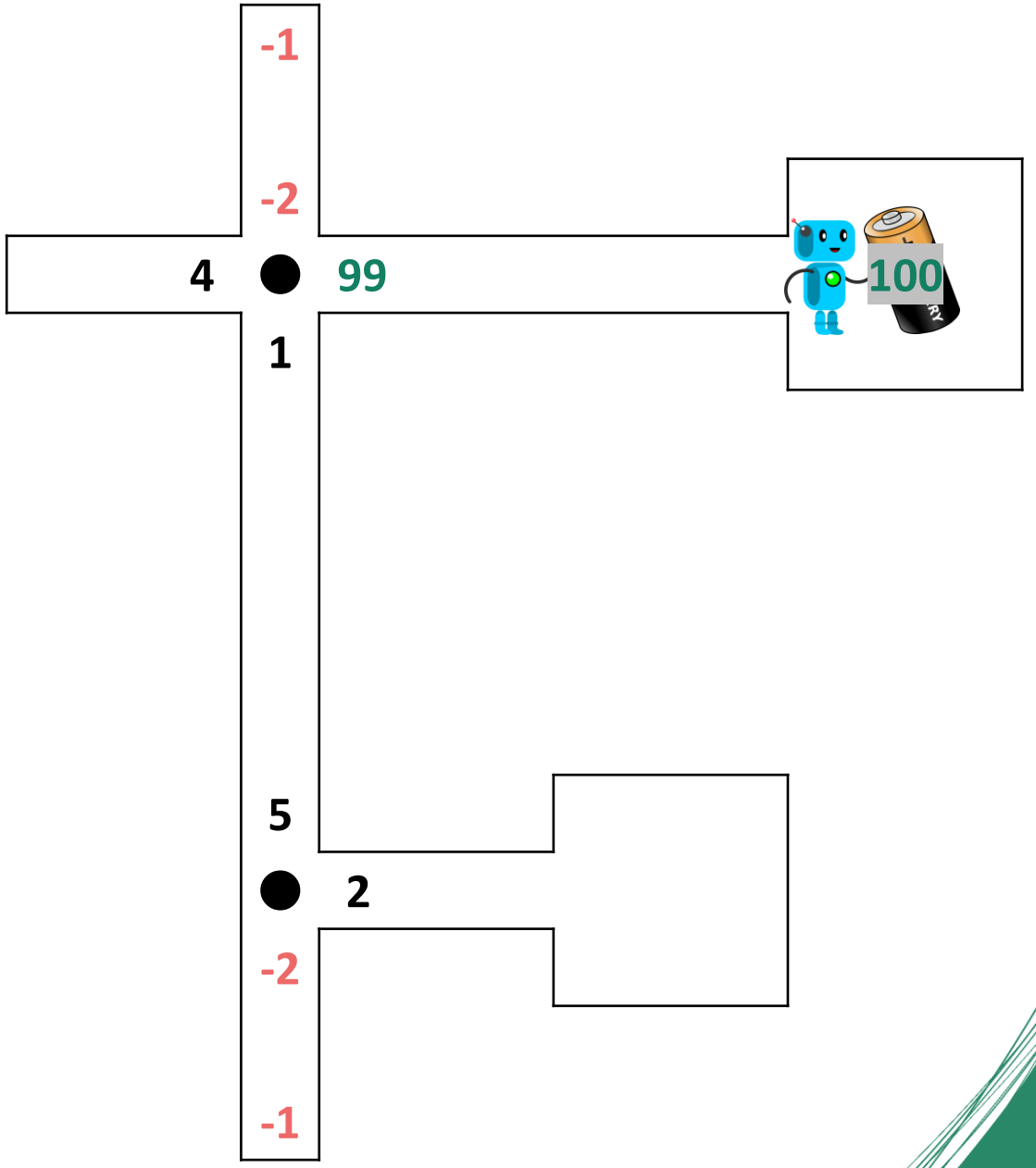
- Das Ziel in diesem Labyrinth hat den Wert **100**





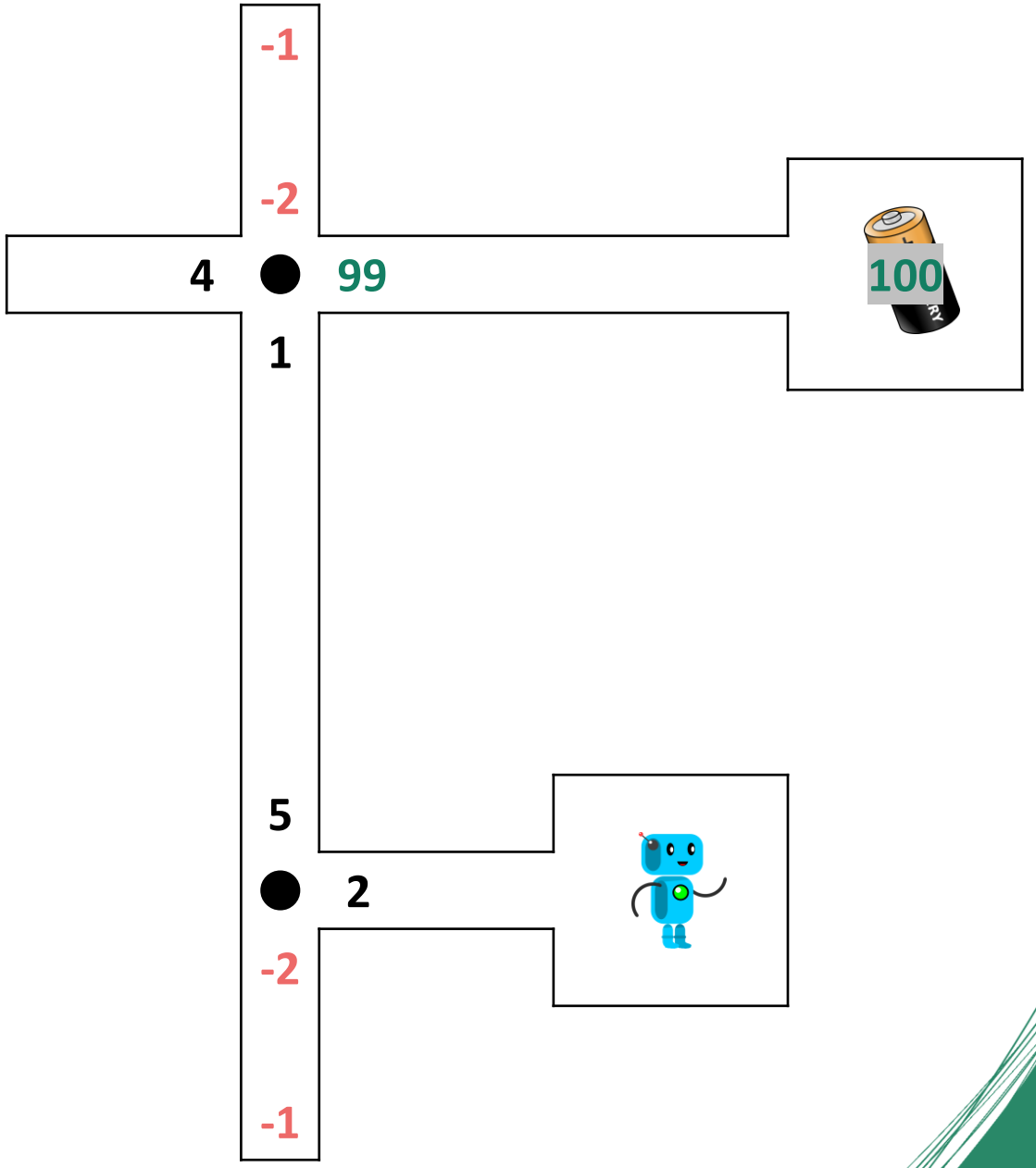
Das Spiel

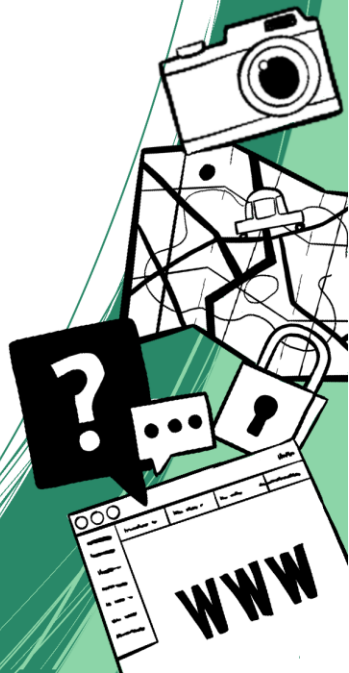
- Das Ziel in diesem Labyrinth hat den Wert **100**
- Vergiss nicht den Weg **upzudaten**!



Das Spiel

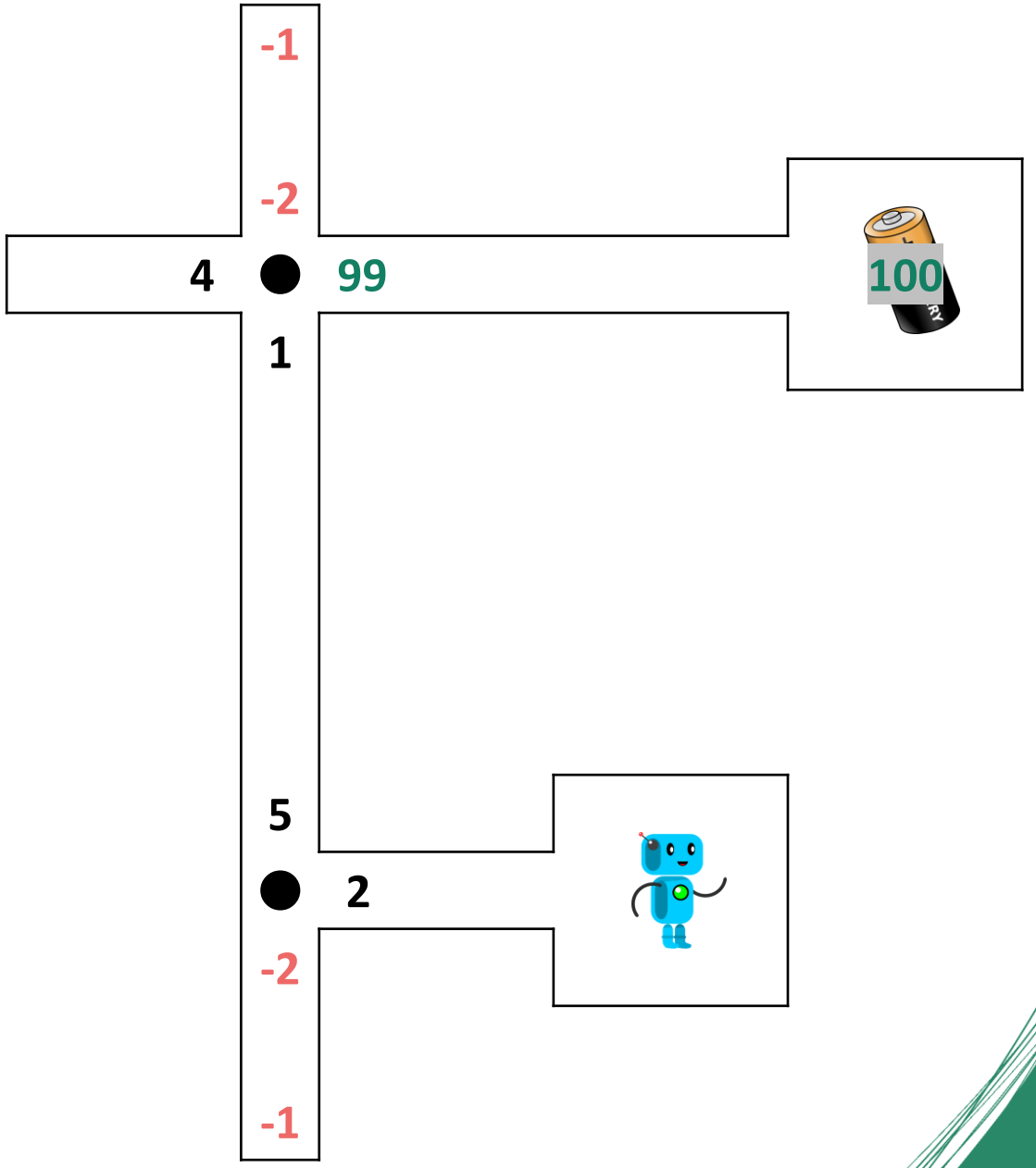
- Das Ziel in diesem Labyrinth hat den Wert **100**
- Vergiss nicht den Weg **upzudaten**!
- Danach kommt der Roboter **zurück zum Start** und beginnt eine **neue Runde**

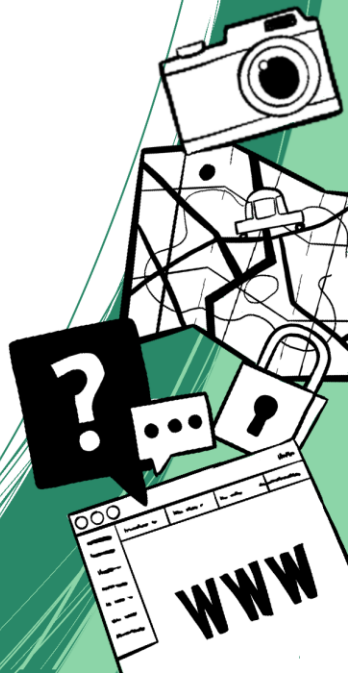




Das Spiel

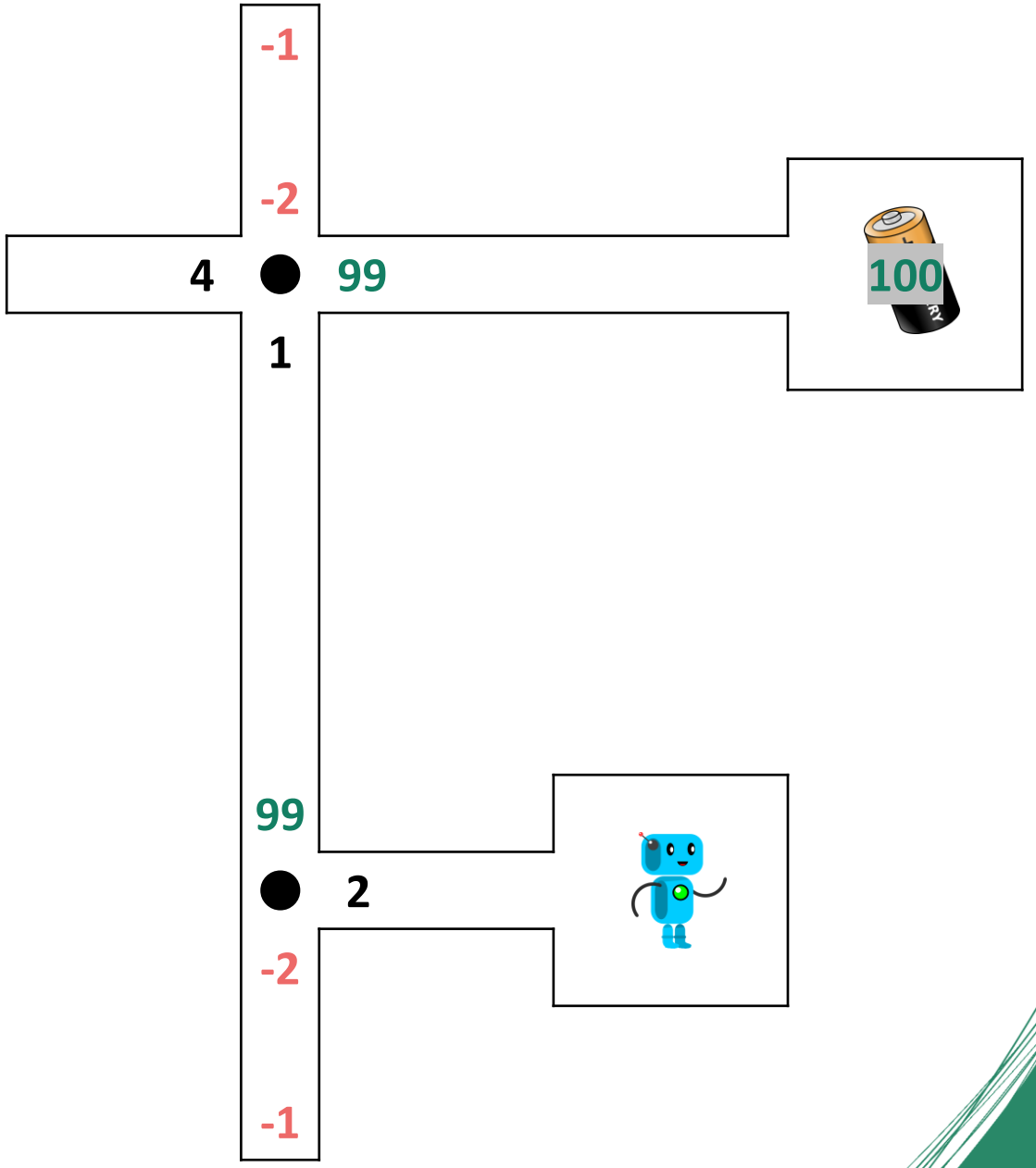
- **Wiederhole** den Vorgang Solange, bis der Roboter **nichts neues mehr dazu lernt** (keine Zahlen mehr schreibt/updated)

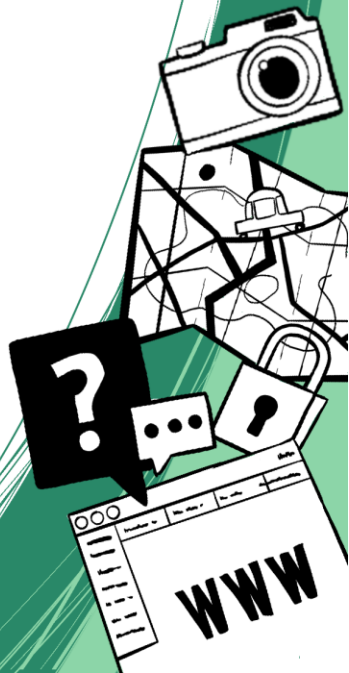




Das Spiel

- Der Roboter hat einen **Weg zum Ziel gelernt!**





ENARIS

Exploration vs Exploitation

Das Problem

- Manche **Aktionen** können zwar **kurzfristig einen positive Effekt**, aber **langfristig einen negative Effekt** haben

Das Problem

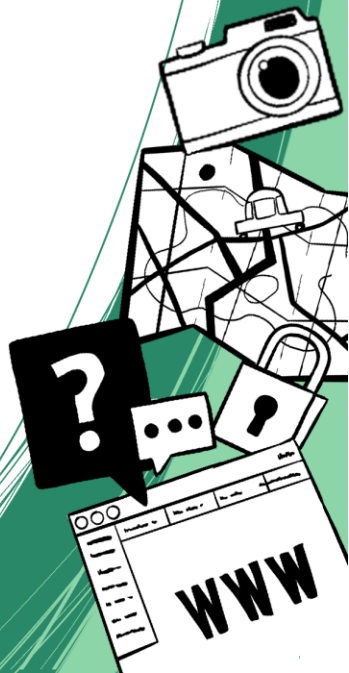
- Manche **Aktionen** können zwar **kurzfristig einen positive Effekt**, aber **langfristig einen negative Effekt** haben
  - Z.B. der Roboter nimmt einen Weg der sich erst viel später als Sackgasse herausstellt.

Das Problem

- Manche **Aktionen** können zwar **kurzfristig einen positive Effekt**, aber **langfristig einen negative Effekt** haben
  - Z.B. der Roboter nimmt einen Weg der sich erst viel später als Sackgasse herausstellt.
- Manche **Aktionen** können kurzfristig nur einen **kleinen positiven Effekt** aber **langfristig einen großen positiven Effekt** haben

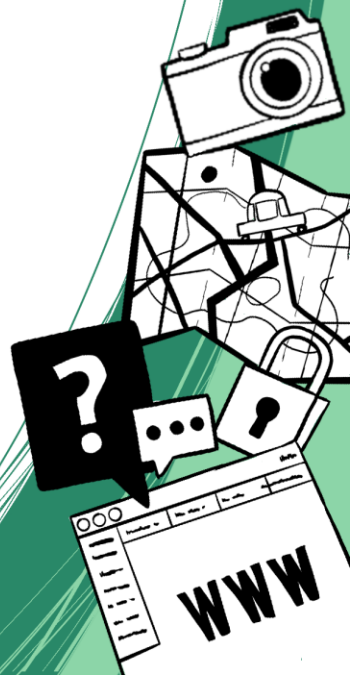
Das Problem

- Manche **Aktionen** können zwar **kurzfristig einen positive Effekt**, aber **langfristig einen negative Effekt** haben
  - Z.B. der Roboter nimmt einen Weg der sich erst viel später als Sackgasse herausstellt.
- Manche **Aktionen** können kurzfristig nur einen **kleinen positiven Effekt** aber **langfristig einen großen positiven Effekt** haben
  - Z.B. der Roboter findet eine Batterie mit einer kleinen Belohnung, es wäre aber eine größere Belohnung entlang des anderen Weges möglich



Die Lösung

- Finde ein **Gleichgewicht** zwischen dem **Verlassen auf bereits gelerntem** (**exploiting**) und dem **Erforschen von neuem** (**exploring**)



Die Lösung

- Finde ein **Gleichgewicht** zwischen dem **Verlassen auf bereits gelerntem** (**exploiting**) und dem **Erforschen von neuem** (**exploring**)
  - Z.B. statt immer den Weg mit dem **höchsten Wert** zu gehen, nimm mit **25% Wahrscheinlichkeit** einen **zufälligen Weg**

Die Lösung

- Finde ein **Gleichgewicht** zwischen dem **Verlassen auf bereits gelerntem** (**exploiting**) und dem **Erforschen von neuem** (**exploring**)
  - Z.B. statt immer den Weg mit dem **höchsten Wert** zu gehen, nimm mit **25% Wahrscheinlichkeit** einen **zufälligen Weg**
  - Diese **Wahrscheinlichkeit** (**exploration rate**) kann auch **dynamisch** sein, sodass sie z.B. im Laufe der Zeit geringer wird