THE INTERNAL MODEL PRINCIPLE OF CONTROL THEORY: A QUICK INTRODUCTION

W.M. Wonham
Dept of Electrical & Computer Engineering
University of Toronto

Nanjing University of Science and Technology (NJUST)
Nanjing, CHINA
2022.09.26

KENNETH J.W. CRAIK
1914-1945

- Cambridge (UK) philosopher and psychologist
- Pioneering “cognitive scientist”

- In WWII worked on human operator in control systems
- Died in Cambridge after road accident

MAN IN THE WORLD

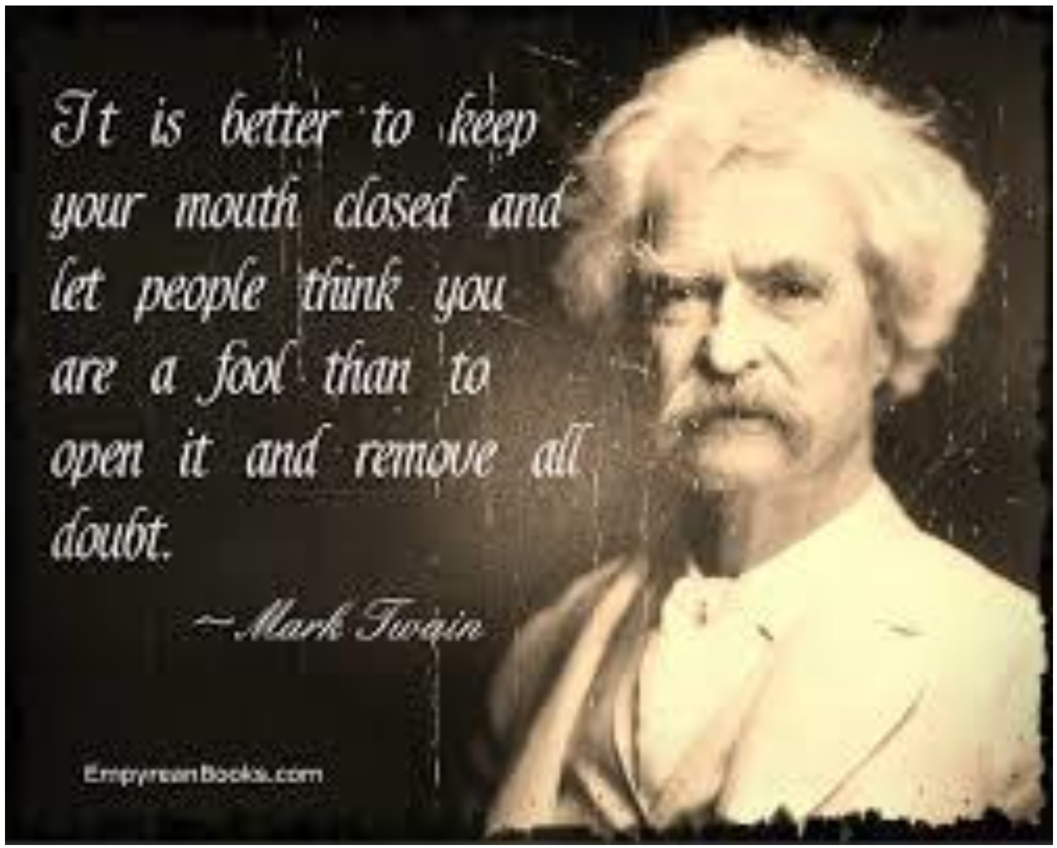
According to Craik,

[O]nly [an] internal model of reality – this working model [in our minds] – enables us to predict events which have not yet occurred in the physical world, a process which saves time, expense, and even life.

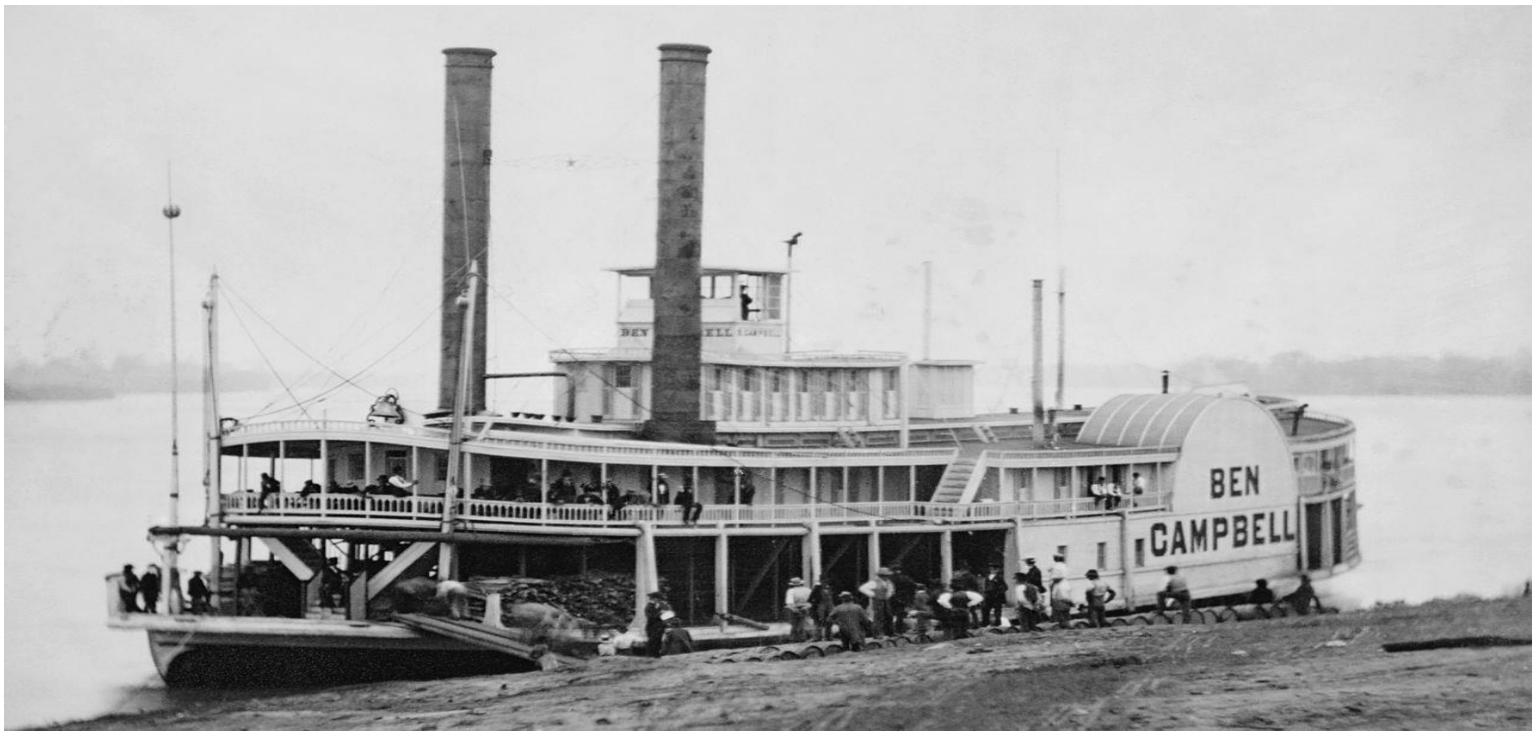
[In other words] the nervous system is viewed as a calculating machine capable of modelling or paralleling external events, and … this process of paralleling is the basic feature of thought and of explanation.

- *The Nature of Explanation*, 1943.

MARK TWAIN (1835-1910)



NAVIGATING THE MISSISSIPPI – 1850s

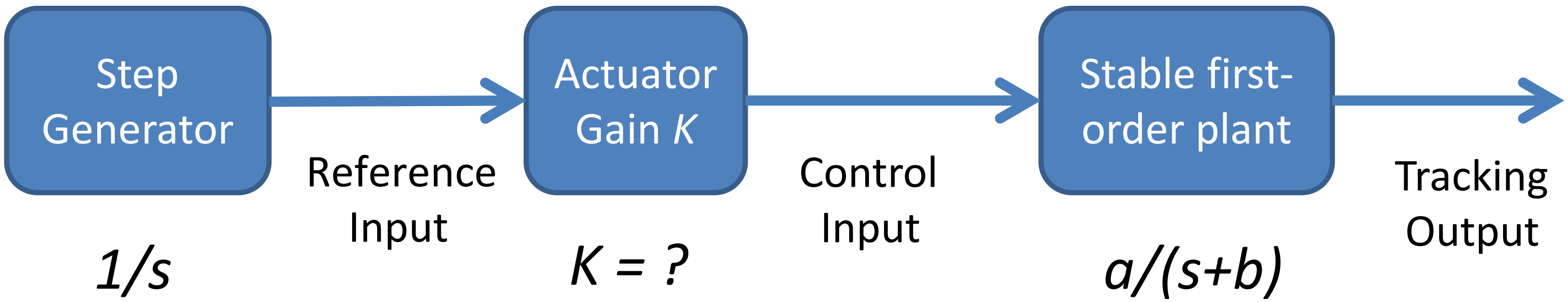


"… You only learn *the* shape of the river; and you learn it with such absolute certainty that you can always steer by the shape that's *in your head,* and never mind the one that's before your eyes."

Mark Twain, *Life on the Mississippi*, 1883

OPEN-LOOP CONTROL

Tracking:

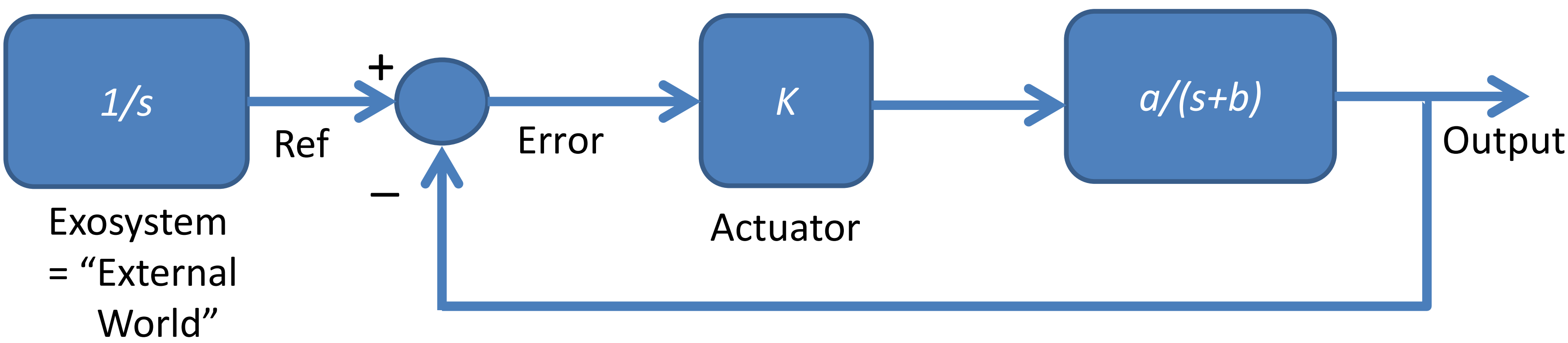


Tracking error := Input – Output → 0 , $t \to \infty$

if and only if

$K = b/a$

ERROR FEEDBACK CONTROL

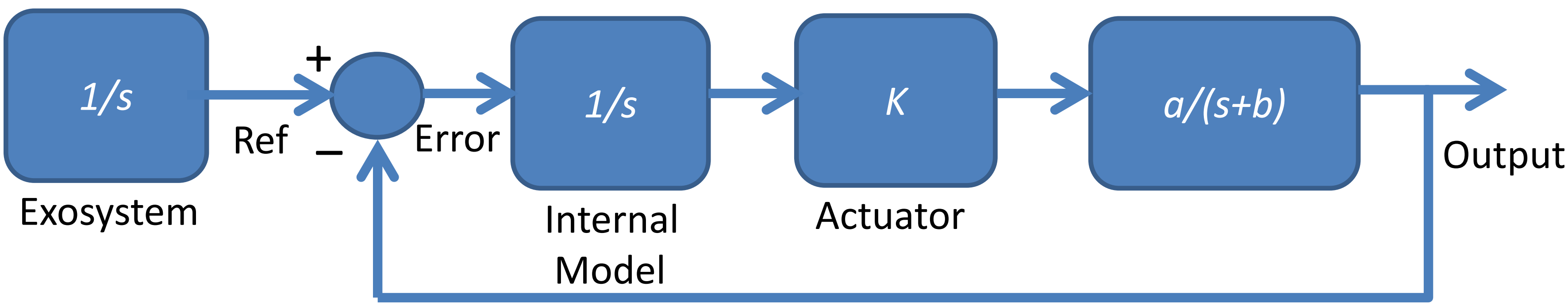


Tracking error → $1/[1+Ka/b]$, $t \to \infty$

$|\textit{Final error}| < \varepsilon$ if and only if $Ka/b > 1/\varepsilon - 1$

Exact value of *K* is not *critical*, but needs to be **large** !

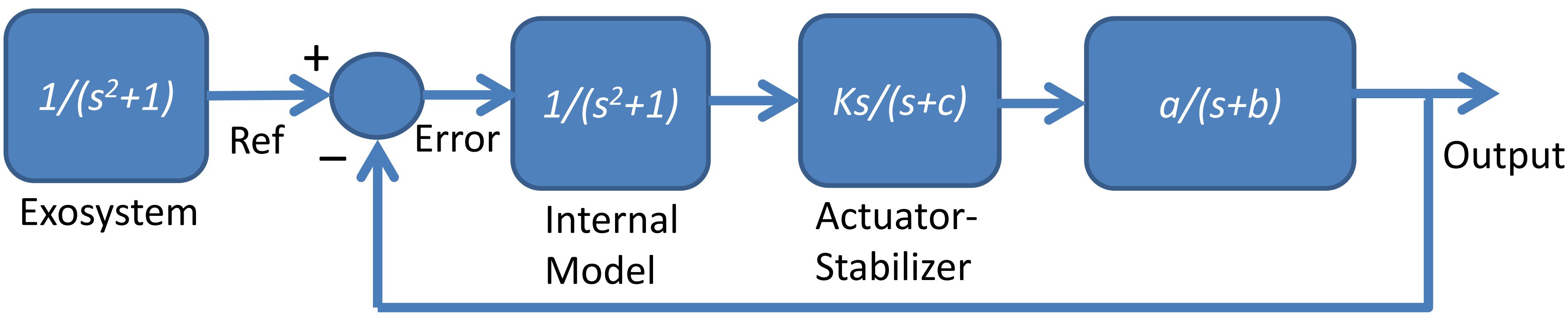
ERROR FEEDBACK WITH (IMPLICIT) INTERNAL MODEL - EXOSYSTEM GENERATES STEPS



Tracking error → *0*, *t*→∞

Final error = 0 for *all K, a, b > 0*.
Exact value of *K* is not critical,
say *Ka/b* ~ *b/2* for good time response.

ERROR FEEDBACK WITH (IMPLICIT) INTERNAL MODEL - EXOSYSTEM GENERATES SINUSOIDS



Tracking error → *0*, $t \to \infty$

Final error = 0 for *all K, a, b ,c > 0,*
for which system is stable.
Exact values of *K* and *c* are not critical.

CONCLUSIONS (PI, PID, PR ~ 1930)

- Error feedback with high loop gain reduces parameter sensitivity and final tracking error.
- Feedback, with (implicit) internal model of exosystem (PI, PID, PR control) reduces final error to exactly zero (perfect tracking), despite (reasonable) parameter perturbations. Requires only moderate average loop gain.
- Price: control complexity, including stabilizer.

ADVANCES (1970s)

- For linear time-invariant multivariable systems with arbitrary (linear) exosystem,

  and for nonlinear systems with step inputs,

  internal model made explicit.

- For such systems, internal model shown to be both necessary and sufficient for structurally stable (***'robust'***) perfect tracking and regulation.

FOR GENERAL SYSTEMS, ASK

- Is error feedback *necessary* for structurally stable perfect regulation (or tracking)?
- Is an internal model *necessary* for structurally stable perfect regulation?
- If "Yes", shouldn't this be true for a *very wide class* of regulator systems, linear or nonlinear?

INTERNAL MODEL PRINCIPLE (IMP)

For a very general class of systems:

1. Error feedback + Perfect regulation

   $\Rightarrow$ Internal Model ?

2. *Structurally stable* perfect regulation

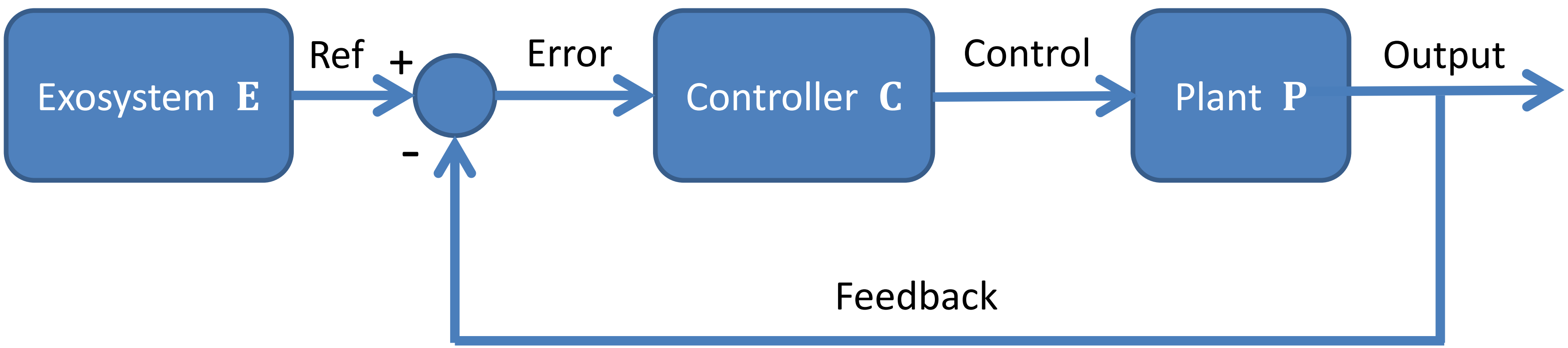
   (regardless of system perturbations)

   $\Rightarrow$ Error feedback + Internal Model ?

GOAL …

- **Establish the IMP**
  in *general* but *rudimentary*
  discrete-time framework,
  using just *ordinary sets and functions*,
  without *any* sophisticated
  technical or geometric assumptions.

TOTAL SYSTEM $\mathbf{S} = \mathbf{E} \times \mathbf{C} \times \mathbf{P}$



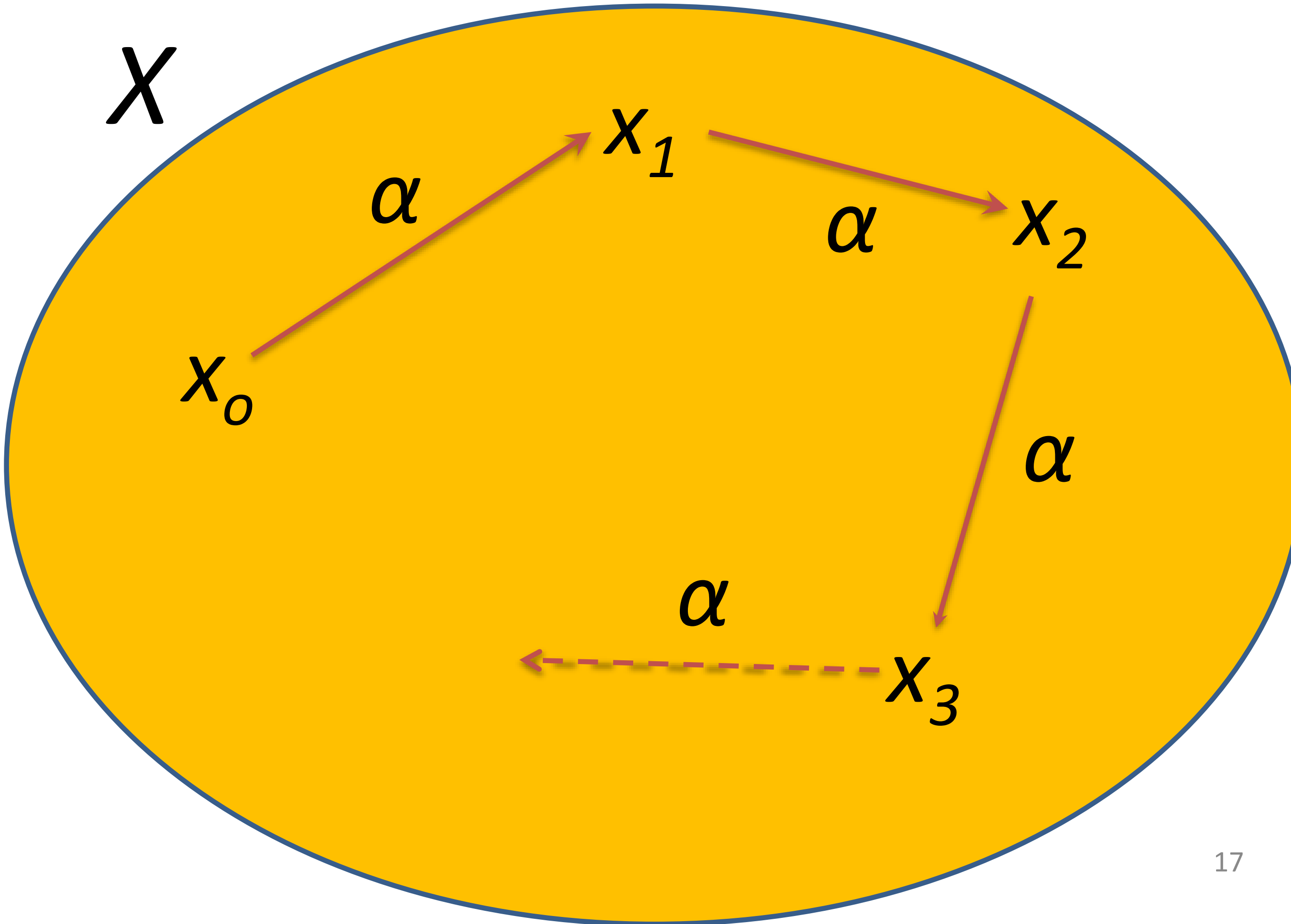
Exosystem = $\mathbf{E}$ = (X_E, α_E) = Dynamic model of *outside world*

Total system = $\mathbf{S}$ = (X,α) = Exosystem × Controller × Plant

WE NEED TO DEFINE …

- (Discrete) dynamical system $\boldsymbol{S}$
- Internal stability
- Feedback structure
- Exosystem detectability

DISCRETE DYNAMICAL SYSTEM **S**: STATE SET *X*, TRANSITION FUNCTION $\alpha: X \rightarrow X$



$x_1=\alpha(x_o)$
$x_2=\alpha(x_1)$
$x_3=\alpha(x_2)$
...

INTERNAL STABILITY

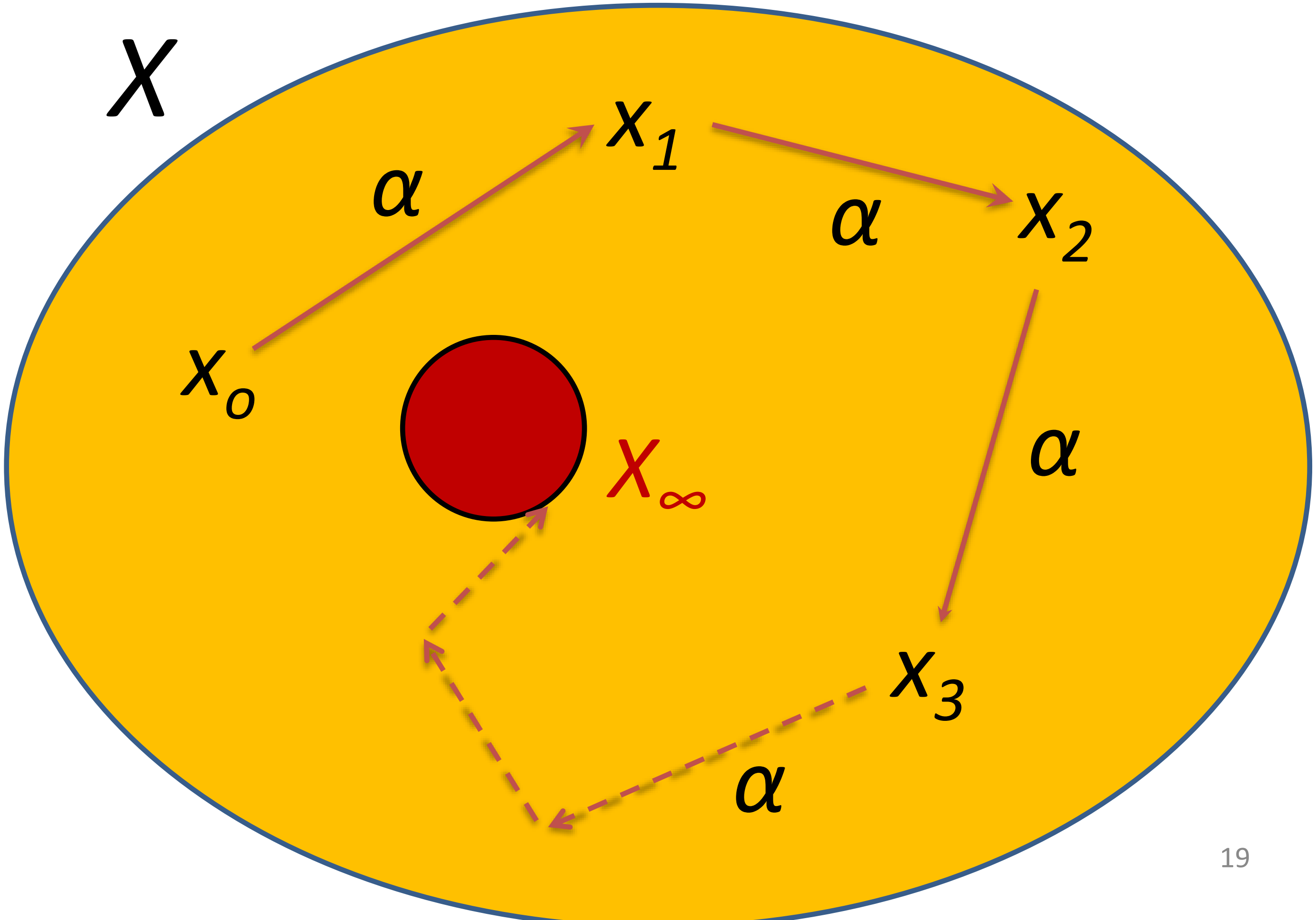
- When driven by the Exosystem $\mathbf{E}$,
  total system $\mathbf{S}$ "approaches steady state"
  and behaves like $\mathbf{E}$ alone.

- E.g. *active* orchestra ($\mathbf{E}$)
  causes *passive audience* ($\mathbf{C} \times \mathbf{P}$)
  to become 'entrained' by dynamic coupling.

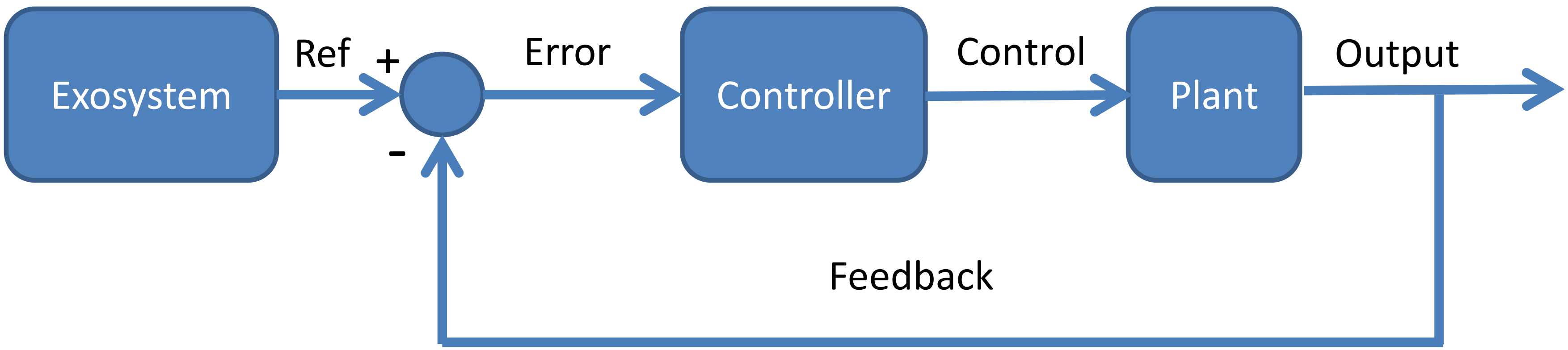
- State $x(t) \to$ *global attractor* $X_\infty =: \widetilde{X}_{\mathbf{E}} \subseteq X$ as $t \to \infty$

INTERNAL STABILITY WITH GLOBAL ATTRACTOR

$x_1=\alpha(x_o)$
$x_2=\alpha(x_1)$
$x_3=\alpha(x_2)$
...



TOTAL SYSTEM $\mathbf{S} = \mathbf{E} \times \mathbf{C} \times \mathbf{P}$ DRIVEN BY EXOSYSTEM $\mathbf{E}$

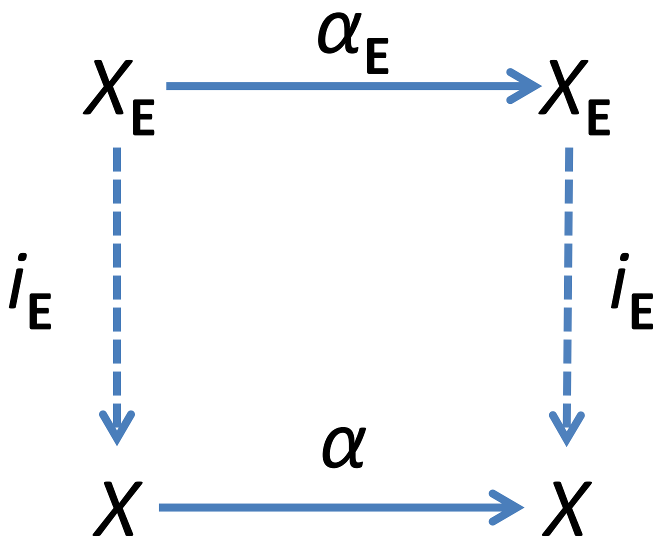


Total system = (Exosystem, Controller, Plant) = $\mathbf{S} = (X,\alpha)$

Exosystem = $\mathbf{E} = (X_{\mathrm{E}}, \alpha_{\mathrm{E}})$

Intuition: If $\mathbf{S}$ is "internally stable" then $\mathbf{E}$ "drives" $\mathbf{S}$ to an "attractor" $X_\infty =: \widetilde{X}_{\mathrm{E}}$ that is α-invariant.

INTERNAL STABILITY – ARROW DIAGRAM



Insertion map i_E injects

$$\mathbf{E} = (X_E, \alpha_E)$$

into $\mathbf{S} = (X, \alpha)$, to create *global attractor*

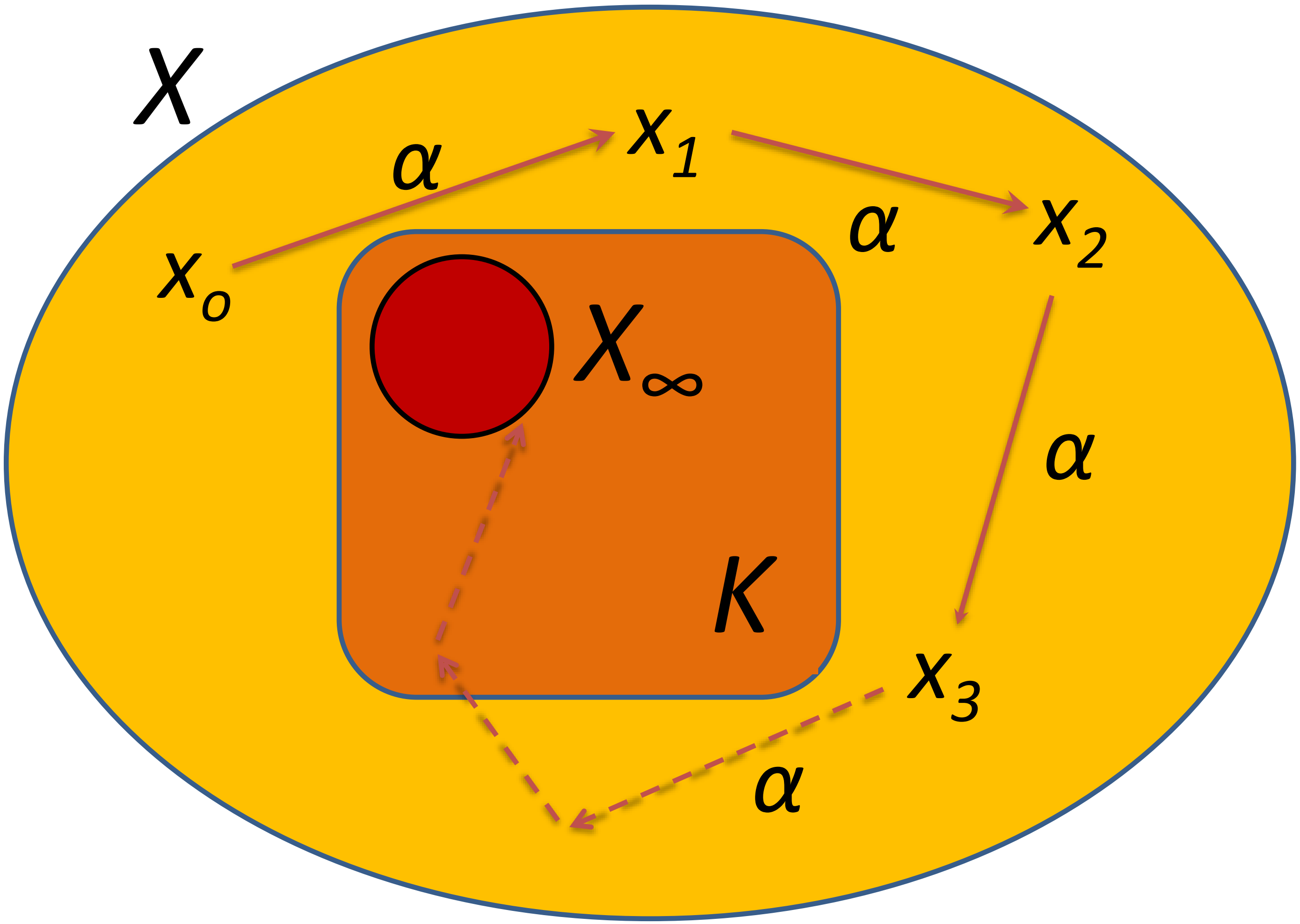
$$i_E(X_E) = \widetilde{X}_E \subseteq X$$

REGULATION (TRACKING): TARGET SUBSET $K \subseteq X$

- *Regulation* (or *tracking*):
  trajectory of $\boldsymbol{S}$
  remains in a suitable *target subset* $K \subseteq X$,
  where "tracking error" is "zero".

- On target subset K,
  $\mathbf{E}$ and $\mathbf{C} \times \mathbf{P}$ track together
  in perfect synchrony

GLOBAL ATTRACTOR $X_\infty = \widetilde{X}_{\mathbf{E}}$ LIES IN TARGET SUBSET K



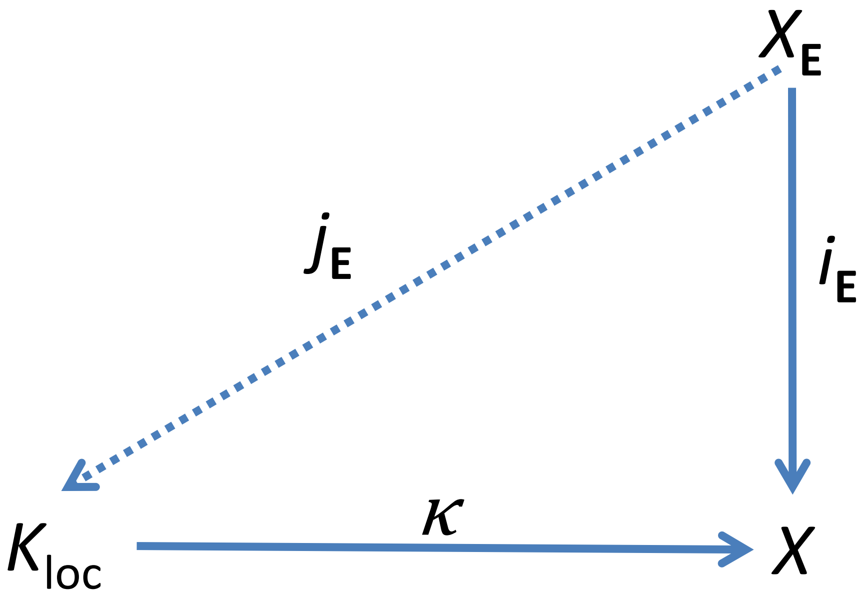
REGULATION CONDITION

- By internal stability, $\widetilde{X}_{\mathbf{E}}$ is a global attractor for the trajectories of $\boldsymbol{S}$.
- For (eventual) regulation we therefore require

$$\widetilde{X}_{\mathbf{E}} \subseteq K$$

- Model $K \subseteq X$ by an insertion $\kappa : K_{\mathrm{loc}} \rightarrow X$.
- Then regulation implies the existence of an injection $j_{\mathbf{E}}$: $X_{\mathbf{E}} \rightarrow K_{\mathrm{loc}}$ with $\kappa \circ j_{\mathbf{E}} = i_{\mathbf{E}}$.

ARROW DIAGRAM FOR REGULATION



$$i_E = \kappa \circ j_E$$

$$\kappa(K_{loc}) = K$$

STATE SET OF CONTROLLER $\mathbf{C}$

- Let *controller* state set

$$X_{\mathbf{C}} = \gamma(X)$$

for suitable surjection (typically, projection)

$$\gamma: X = X_{\mathbf{E}} \times X_{\mathbf{C}} \times X_{\mathbf{P}} \to X_{\mathbf{C}}$$

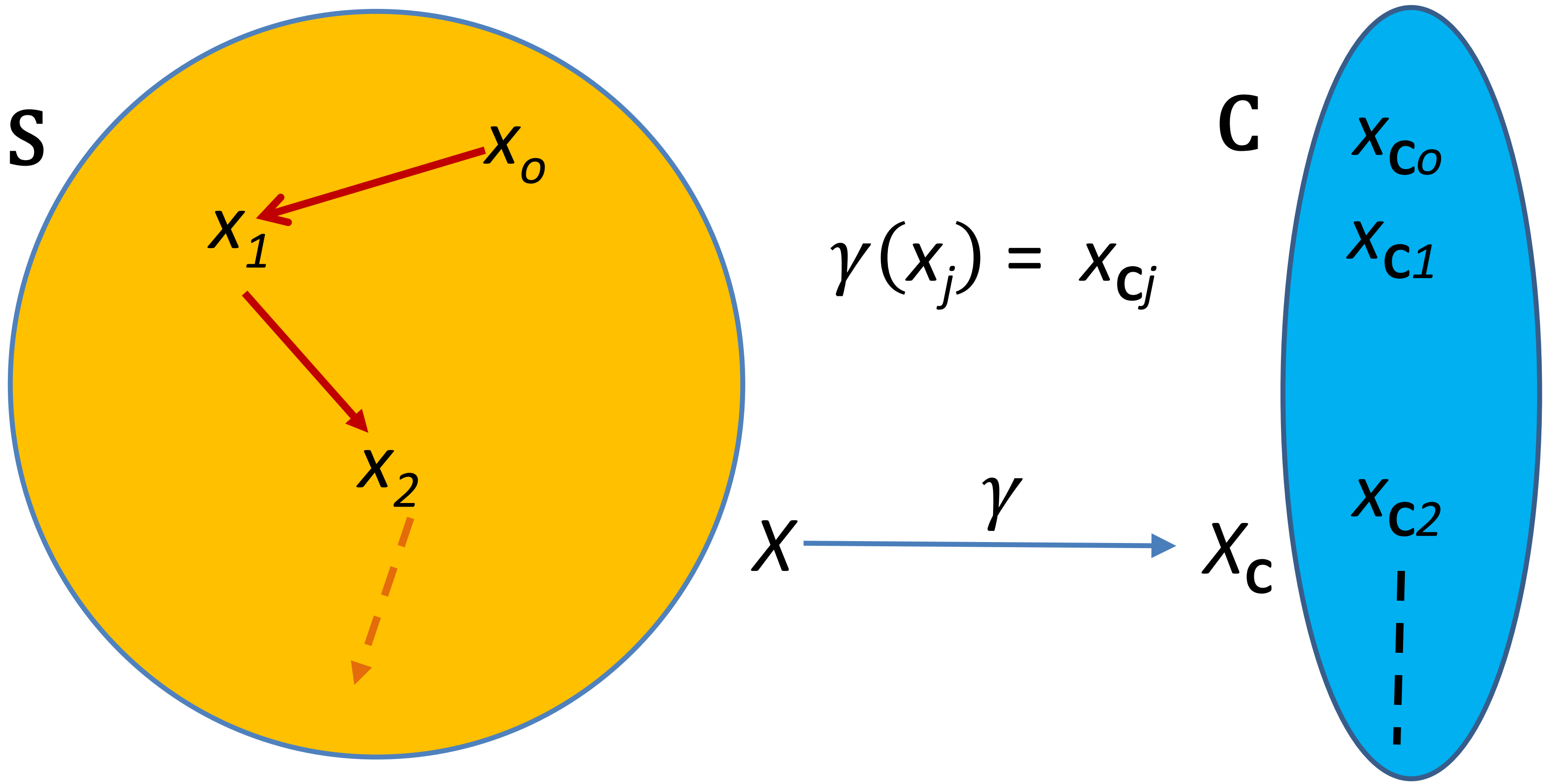
- So we have

Total state (of $\mathbf{S}$) = $x \in X$

$\Rightarrow$ Controller state x_c (of $\mathbf{C}$) = $\gamma(x) \in X_{\mathbf{C}}$

SYSTEM 'OBSERVED' BY CONTROLLER

- Transition map α: $X \rightarrow X$, Control map γ: $X \rightarrow X_{\mathbf{c}}$



FEEDBACK STRUCTURE

- Controller $\mathbf{C}$ is *externally* driven
  only when state of $\mathbf{S}$ *deviates* from regulation
  target set *K*.
- Dynamics of $\mathbf{C}$ are *autonomous:*
  as long as $x \in K$, "tracking remains perfect."
- Suppose $x \in K$, so controller state is $x_{\mathbf{c}} = \gamma(x)$.
  By feedback, 'next' controller state $x_{\mathbf{c}}' = \gamma(\alpha(x))$
  depends only on $x_{\mathbf{c}} = \gamma(x)$, i.e. for $x \in K$, $\gamma \circ \alpha(x)$
  can be computed from $\gamma(x)$. Technically,
  $$\ker(\gamma | K) \;\leq\; \ker(\gamma \circ \alpha | K)$$
- [But needn't be true that *K* is α-invariant!]

DETECTABILITY OF EXOSYSTEM ON GLOBAL ATTRACTOR

- Consider motion of $\mathbf{S}$ on global attractor $i_{\mathbf{E}}(X_{\mathbf{E}}) =: \widetilde{X}_{\mathbf{E}}$. By definition, $\alpha(\widetilde{X}_{\mathbf{E}}) \subseteq \widetilde{X}_{\mathbf{E}}$.
- If $x(0) = x_{\mathrm{o}} \in \widetilde{X}_{\mathbf{E}}$, then $x(t) = \alpha^t(x_{\mathrm{o}}) \in \widetilde{X}_{\mathbf{E}}$, $t = 0,1,2,\ldots$
- Assume controller could (eventually) identify x_{o} from observations $x_{\mathbf{C}}(t) = \gamma(x(t)), t = 0,1,2,\ldots$
- This simply means:

  $\widetilde{X}_{\mathbf{E}}$ is *detectable* by $\mathbf{S} := (X,\alpha,\gamma)$

  In other words

  $\widetilde{\mathbf{S}}_{\mathbf{E}} := (\widetilde{X}_{\mathbf{E}}, \alpha|\widetilde{X}_{\mathbf{E}}, \gamma|\widetilde{X}_{\mathbf{E}})$ is *observable*.

THEOREM: INTERNAL MODEL PRINCIPLE

- With $\widetilde{X}_{\mathbf{E}}$ as defined above, write
  $\widetilde{\alpha}_{\mathbf{E}} := \alpha|\widetilde{X}_{\mathbf{E}}$, $\widetilde{\gamma}_{\mathbf{E}} := \gamma|\widetilde{X}_{\mathbf{E}}$

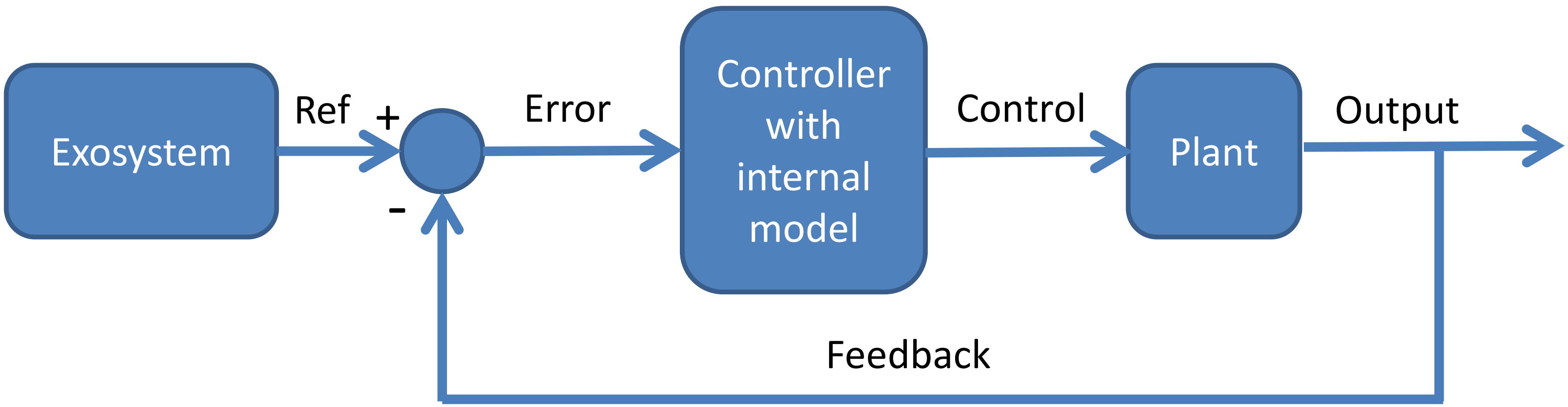
- **Theorem**: Assume that $\boldsymbol{S}$ satisfies *internal stability*, *regulation*, *feedback structure*, and *exosystem detectability*. Then

  1. there exists a *unique* mapping $\alpha_{\mathbf{C}}: X_{\mathbf{C}} \to X_{\mathbf{C}}$ determined by $\alpha_{\mathbf{C}} \circ \gamma_{\mathbf{K}} = \gamma \circ \alpha|K$
  2. $\alpha_{\mathbf{C}} \circ \widetilde{\gamma}_{\mathbf{E}} = \widetilde{\gamma}_{\mathbf{E}} \circ \widetilde{\alpha}_{\mathbf{E}}$
  3. $\widetilde{\gamma}_{\mathbf{E}}$ is injective

INTERPRETATION

- **Statement 1** defines the controller's dynamics, as *autonomous* under the condition of regulation.
- **Statement 2** identifies these controller dynamics as a *copy of the dynamics of* $\mathbf{E}$ on the global attractor (i.e. *exosystem* dynamics).
- **Statement 3** asserts that this copy is *faithful*, namely incorporates fully the exosystem dynamics.

TOTAL SYSTEM $\mathbf{S} = \mathbf{E} \times \mathbf{C} \times \mathbf{P}$ DRIVEN BY EXOSYSTEM $\mathbf{E}$ ACHIEVES PERFECT TRACKING

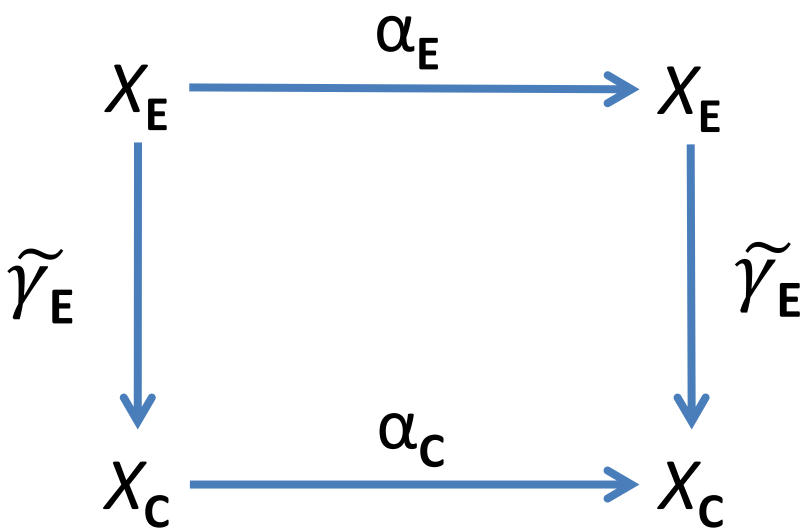


Total system = (Exosystem, Controller, Plant) = $\mathbf{S} = (X,\alpha)$

Exosystem = $\mathbf{E} = (X_E, \alpha_E)$

With internal model in $\mathbf{C}$, $\mathbf{E}$ "drives" $\mathbf{S}$ to $\tilde{X}_E \subseteq K$, where tracking remains perfect, namely Error ≡ 0.

FINAL RESULT



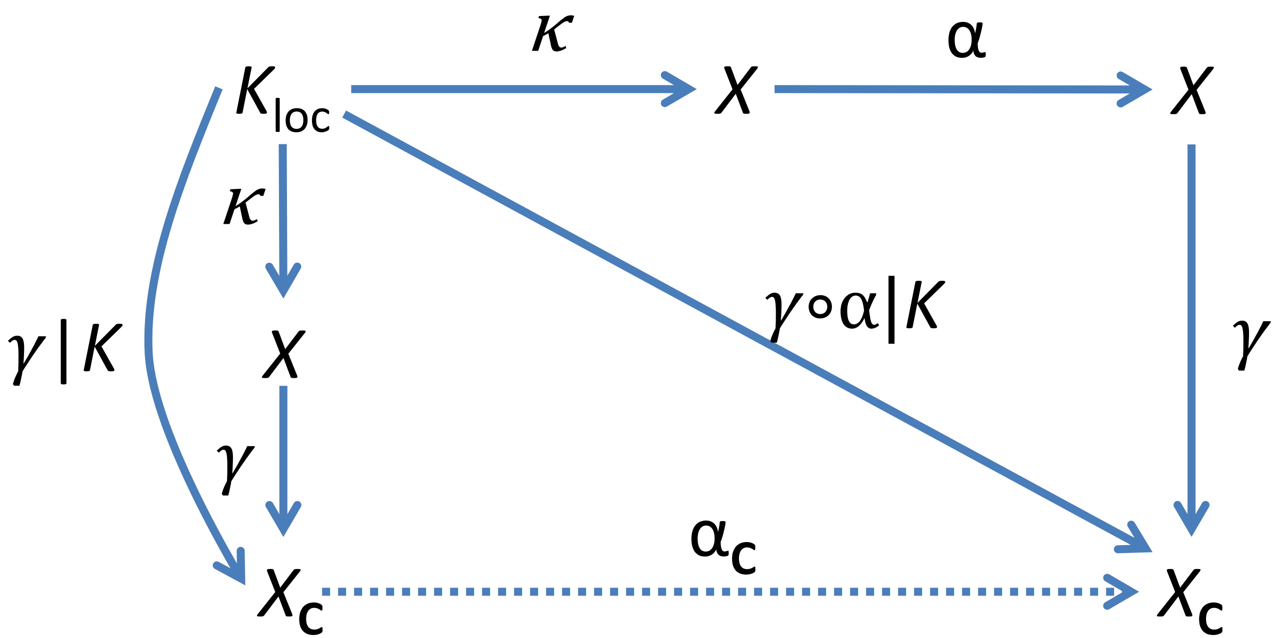
Internal model of exosystem in the controller is *faithfully represented* by $(\widetilde{\gamma}_E(X_E),\ \alpha_C|\widetilde{\gamma}_E(X_E))$

PROOF OF THEOREM (1)

- **Statement 1**. *Derive* controller dynamics $\alpha_{\mathbf{c}}$

  Let $x_{\mathbf{c}} \in X_{\mathbf{c}}$.
  By $\gamma(K) = X_{\mathbf{c}}$, there is $x \in K$ (so $x \in X$) with $\gamma(x) = x_{\mathbf{c}}$.
  Define $\alpha_{\mathbf{c}}(x_{\mathbf{c}}) := \gamma \circ \alpha(x)$.
  The definition is unambiguous, since
  $x' \in K \ \ \& \ \gamma(x') = x_{\mathbf{c}} \ \Rightarrow \ x' \equiv x \ (mod \ \ker(\gamma))$,
  namely (as $x, x' \in K$),
  $x' \equiv x \ (mod \ \ker(\gamma | K))$.
  By feedback, $x' \equiv x \ (mod \ \ker(\gamma \circ \alpha | K))$, so
  $\gamma \circ \alpha(x') = \gamma \circ \alpha(x)$.

STATEMENT 1



Existence of autonomous control dynamics, operating while regulation is satisfied.

PROOF OF THEOREM (2)

- **Statement 2**. Controller dynamics *imitates* $\mathbf{E}$ on K

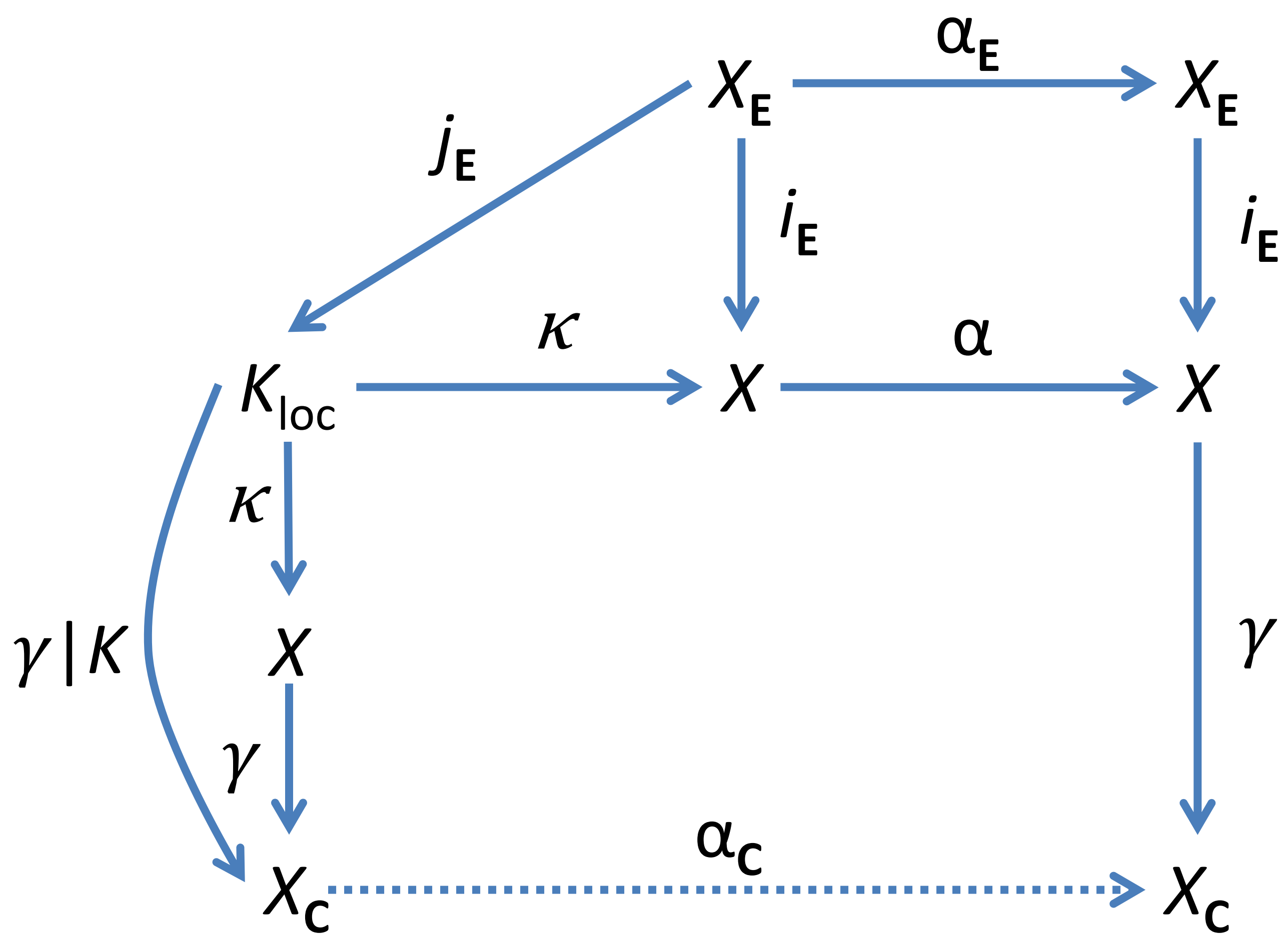
  Let $x \in \widetilde{X}_{\mathbf{E}}$.
  Since $\widetilde{X}_{\mathbf{E}} \subseteq K$, from Statement 1 we get
  $$\alpha_{\mathbf{C}} \circ \widetilde{\gamma}_{\mathbf{E}}(x) = \gamma \circ \alpha(x).$$
  But
  $$\alpha(x) = \widetilde{\alpha}_{\mathbf{E}}(x) \text{ and } \alpha(x) \in \widetilde{X}_{\mathbf{E}},$$
  so
  $$\alpha_{\mathbf{C}} \circ \widetilde{\gamma}_{\mathbf{E}}(x) = \widetilde{\gamma}_{\mathbf{E}} \circ \widetilde{\alpha}_{\mathbf{E}}(x)$$
  as claimed.

PROOF OF THEOREM (3)

- **Statement 3**. Imitation of $\mathbf{E}$ by $\mathbf{C}$ is *faithful*

  Let $\omega \in \mathcal{E}(X)$ be the observer for (γ,α), and
  $\tilde{\omega}_{\mathbf{E}} := \omega \,|\, \tilde{X}_{\mathbf{E}}$ its restriction to $\tilde{X}_{\mathbf{E}}$.
  By observer theory
  $\tilde{\omega}_{\mathbf{E}} = \sup\{\omega' \in \mathcal{E}(\tilde{X}_{\mathbf{E}}) \mid \omega' \leq \ker(\tilde{\gamma}_{\mathbf{E}}) \wedge (\omega' \bullet \tilde{\alpha}_{\mathbf{E}})\}$
  $= \perp$ (using detectability of $\tilde{X}_{\mathbf{E}}$).
  Also, feedback with restrictions to $\tilde{X}_{\mathbf{E}}$ implies
  $\ker(\tilde{\gamma}_{\mathbf{E}}) \leq \ker(\tilde{\gamma}_{\mathbf{E}} \circ \tilde{\alpha}_{\mathbf{E}}) = (\ker(\tilde{\gamma}_{\mathbf{E}})) \bullet \tilde{\alpha}_{\mathbf{E}}$.
  Therefore $\ker(\tilde{\gamma}_{\mathbf{E}}) \leq \tilde{\omega}_{\mathbf{E}} = \perp$,
  namely $\tilde{\gamma}_{\mathbf{E}}$ is injective, as claimed.

THEOREM - IMP



$\widetilde{\gamma}_E := \gamma | \widetilde{X}_E = \gamma \circ i_E = \gamma \circ (\kappa \circ j_E) = (\gamma \circ \kappa) \circ j_E = (\gamma | K) \circ j_E$.

Therefore $\alpha_C \circ \widetilde{\gamma}_E = \widetilde{\gamma}_E \circ \alpha_E$, as displayed before.

WORKING IN **ARROWS**

- IMP has been developed solely in the language of *arrows*!
- Just sets and functions - no sophisticated or difficult mathematics like differential equations or functional analysis!
- These arrows can be generalized far beyond sets and functions – to things called “topoi”!
- Will this be the cybernetic language of the future?

INTERNAL MODEL PRINCIPLE (IMP)

For a very general class of systems:

1. Error feedback + Perfect regulation

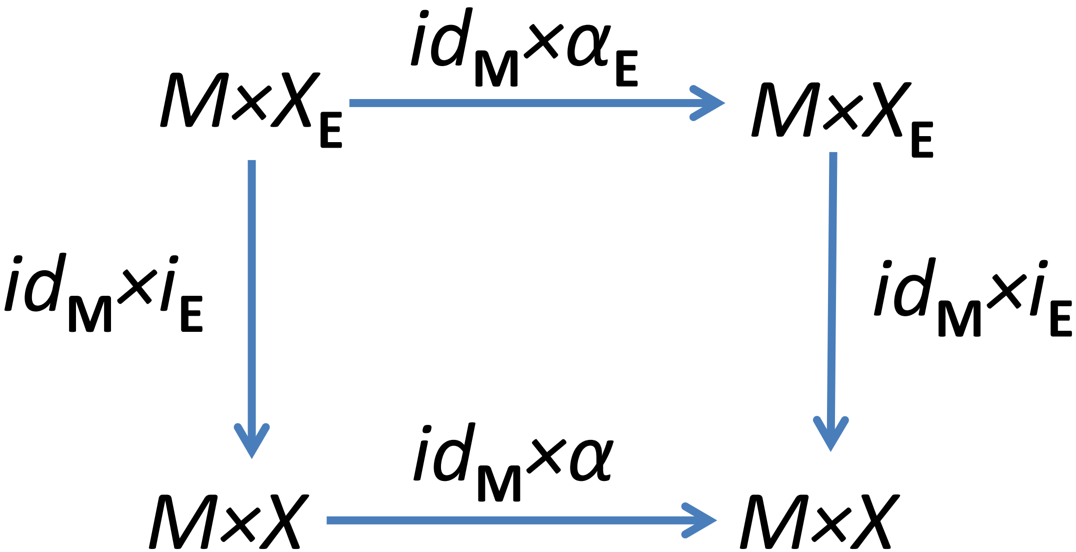
   $\Rightarrow$ Internal Model ✓

2. *Structurally stable* perfect regulation

   (i.e. regulation *regardless* of system perturbations)

   $\Rightarrow$ Error feedback + Internal Model ?

PARAMETRIZED INTERNAL STABILITY



M = set of static system parameter elements μ

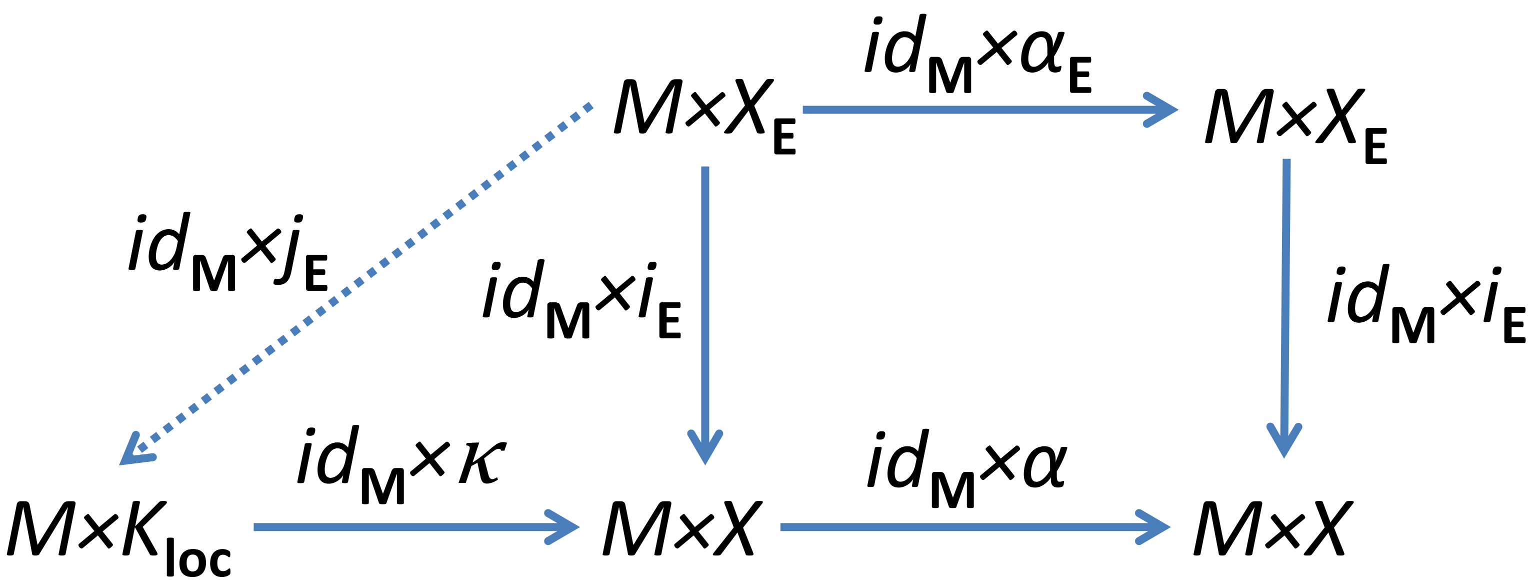
Global attractor $i_E(\{\mu\} \times X_E) \subseteq X$

now depends on the system parameter $\mu \in M$!

For all $\mu \in M$, $x_E \in X_E$, there holds

$$\alpha(\mu, i_E(\mu, x_E)) = i_E(\mu, \alpha_E(x_E))$$

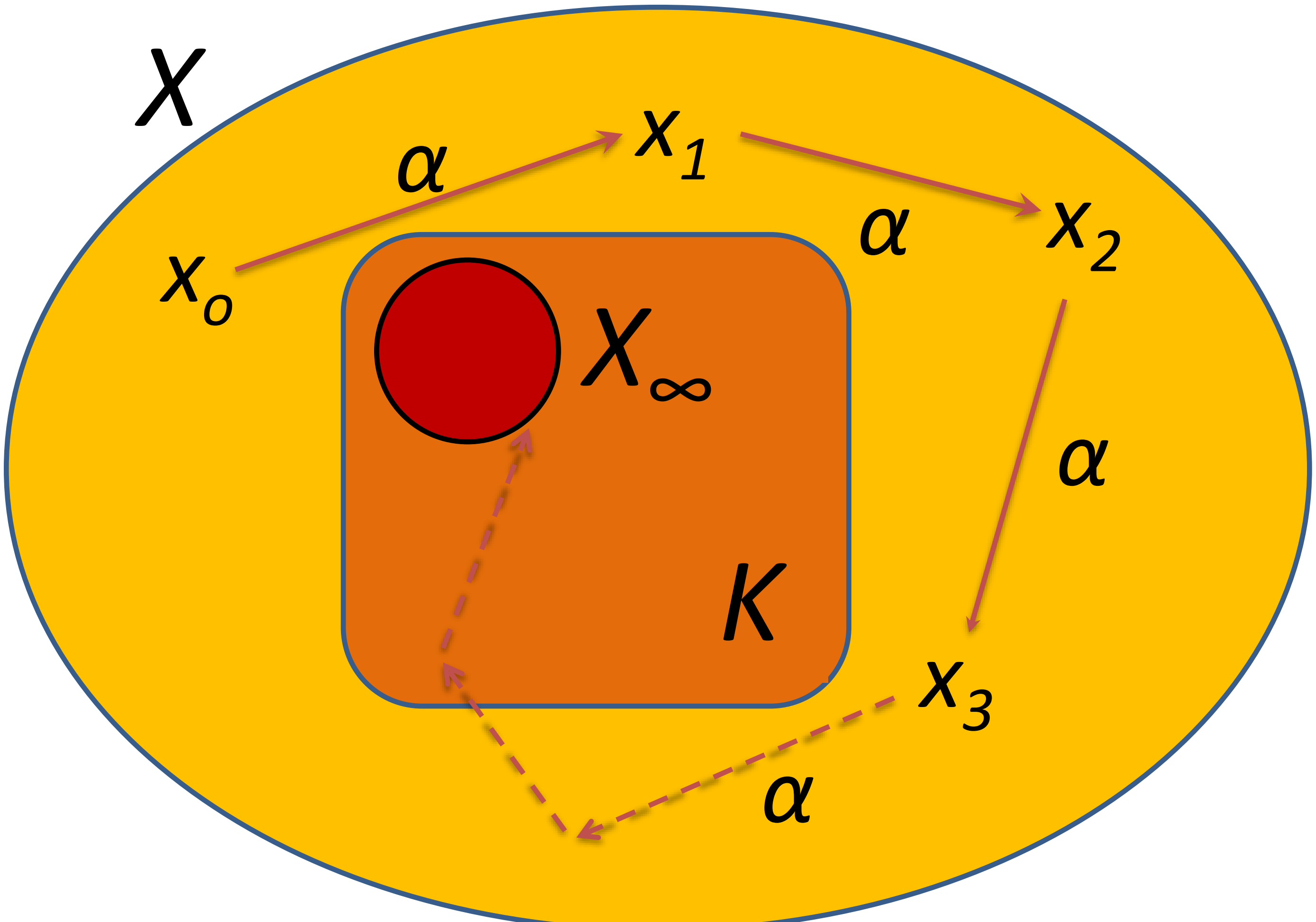
STRUCTURALLY STABLE REGULATION



Property assumed:

For all $\mu \in M$ and $x_E \in X_E$, have $i_E(\mu, x_E) \subseteq K \subseteq X$, where K is fixed and independent of μ. Thus for suitable $j_E : M \times X_E \rightarrow K_{loc}$, $i_E(\mu, x_E) \equiv \kappa(j_E(\mu, x_E))$.

GLOBAL ATTRACTOR $X_\infty = \widetilde{X}_{\mathbf{E}}$ LIES IN TARGET SUBSET K



DEDUCING FEEDBACK STRUCTURE

HARMLESS TECHNICAL ASSUMPTIONS:

- Regulation target subset K constrains $(x_{\mathbf{E}}, x_{\mathbf{P}})$, but not $x_{\mathbf{C}}$

- Parameter product structure

  $M = M_{\mathbf{E}} \times M_{\mathbf{P}} \times M_{\mathbf{C}}$

  (with component sets pairwise disjoint)

  $\mu = (\mu_{\mathbf{E}},\ \mu_{\mathbf{P}},\ \mu_{\mathbf{C}})$

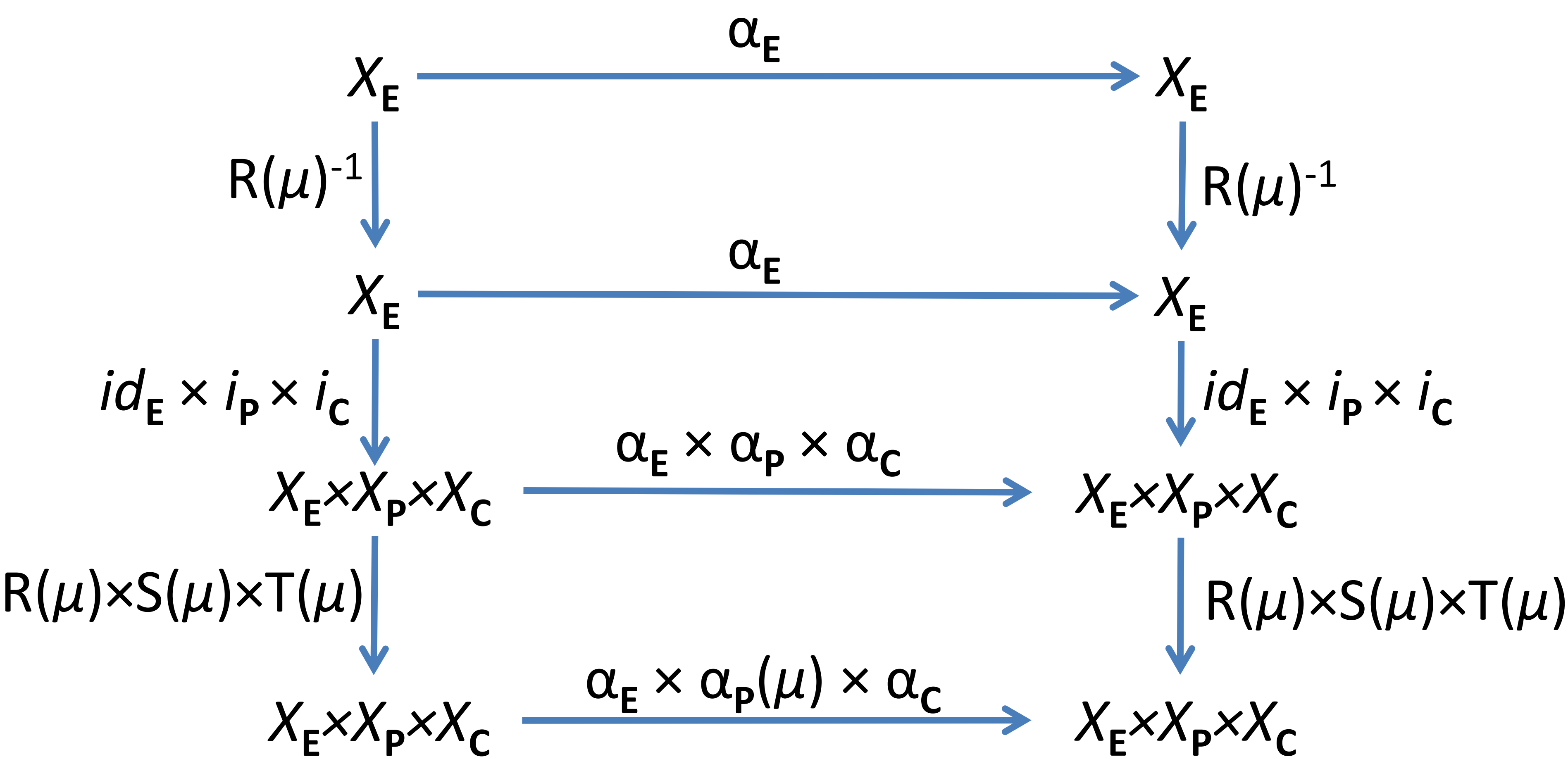
  (independent components)

ADMISSIBLE TRANSFORMATIONS OF $\mathbf{E}$, $\mathbf{C}$, and $\mathbf{P}$

$$
\begin{array}{ccc}
X_{\mathrm{E}} & \xrightarrow{\alpha_{\mathrm{E}}} & X_{\mathrm{E}} \\
{\scriptstyle \mathrm{R}(\mu)}\downarrow & & \downarrow{\scriptstyle \mathrm{R}(\mu)} \\
X_{\mathrm{E}} & \xrightarrow{\alpha_{\mathrm{E}}} & X_{\mathrm{E}}
\end{array}
\qquad\qquad
\begin{array}{ccc}
X & \xrightarrow{\alpha_{\mathrm{C}}} & X_{\mathrm{C}} \\
{\scriptstyle (\mathrm{R}\times\mathrm{S}\times\mathrm{T})(\mu)}\downarrow & & \downarrow{\scriptstyle \mathrm{T}(\mu)} \\
X & \xrightarrow{\alpha_{\mathrm{C}}} & X_{\mathrm{C}}
\end{array}
$$

$$
\begin{array}{ccc}
X & \xrightarrow{\alpha_{\mathrm{P}}} & X_{\mathrm{P}} \\
{\scriptstyle (\mathrm{R}\times\mathrm{S}\times\mathrm{T})(\mu)}\downarrow & & \downarrow{\scriptstyle \mathrm{S}(\mu)} \\
X & \xrightarrow{\alpha_{\mathrm{P}}(\mu)} & X_{\mathrm{P}}
\end{array}
$$

PERTURBATION MODEL



RICH PARAMETER PERTURBATION

- From arrow diagrams,

  $\alpha_C\,[R(\mu_E)(x_E), S(\mu_P)\circ i_P(x_E), T(\mu_C)\circ i_C\,(x_E)] = T(\mu_C)\circ i_C\circ\alpha_E\,(x_E)$

- Crucial assumption - **Rich Parameter Perturbation:**

  For each fixed x_E,

  as μ_E varies through M_E and μ_P varies through M_P,

  $R(\mu_E)(x_E)$ *varies through* X_E and $S(\mu_P)\circ i_P(x_E)$ *varies through* X_P

- Therefore

  $\alpha_C\,[R(\mu_E)(x_E), S(\mu_P)\circ i_P(x_E), T(\mu_C)\circ i_C\,(x_E)]$

  depends only on $T(\mu_C)\circ i_C\,(x_E)$

CONCLUSION

- For each fixed parameter value μ,
  the system $\mathbf{S}$ has *feedback structure*
  on attractor $\widetilde{X}_{\mathbf{E}}(\mu)$

- So for every μ, controller $\mathbf{C}$ is *autonomous*
  when regulation is perfect

- As before, deduce:

  $\mathbf{C}$ contains an Internal Model of $\mathbf{E}$

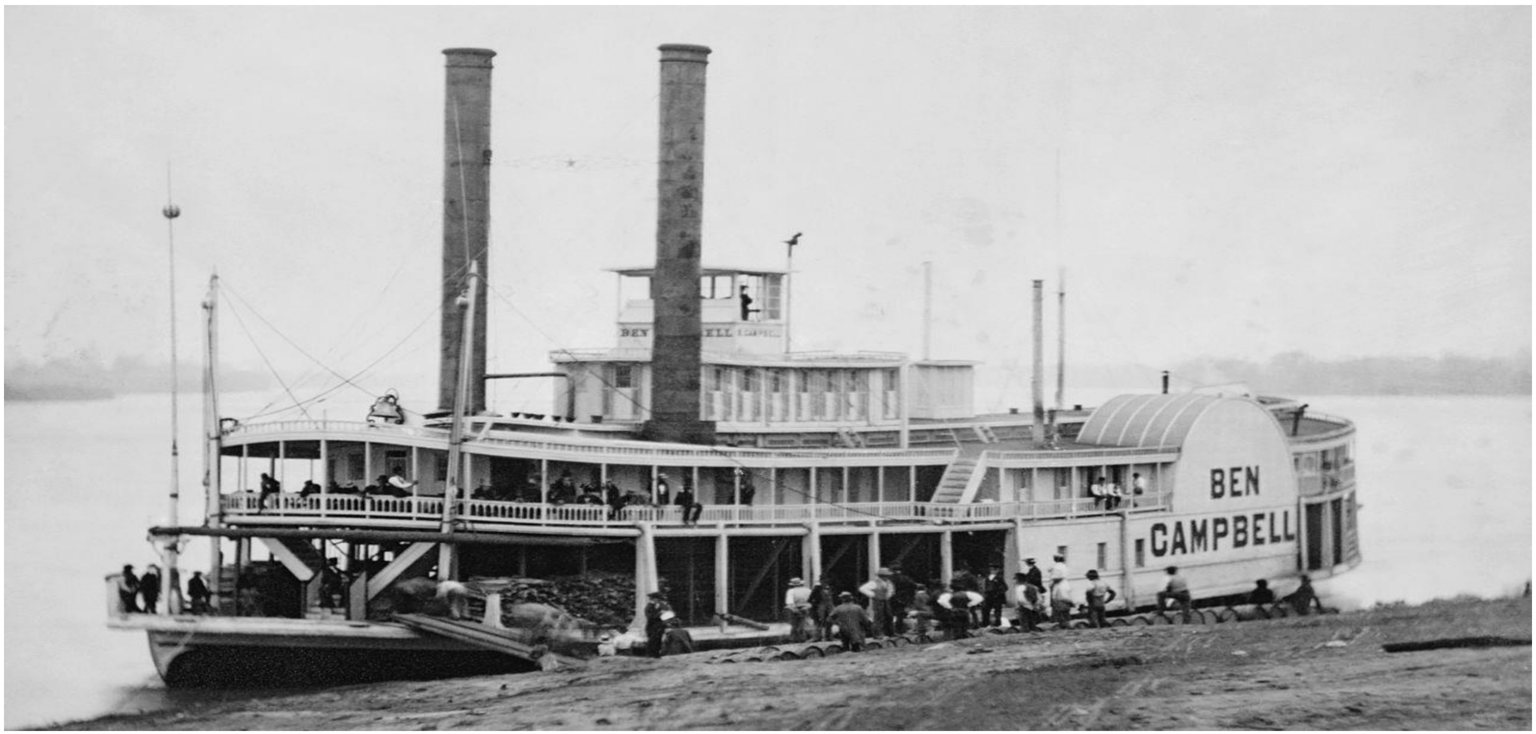
INTERNAL MODEL PRINCIPLE (IMP)

For a very general class of systems:

1. Error feedback + Perfect regulation
   $\Rightarrow$ Internal Model ✓

2. *Structurally stable* (or *“robust”*) perfect regulation
   (i.e. regulation regardless of system perturbations)
   $\Rightarrow$ Error feedback + Internal Model ✓

BACK TO THE MISSISSIPPI – 1850s



"… You only learn *the* shape of the river; and you learn it with such absolute certainty that you can always steer by the shape that's *in your head,* and never mind the one that's before your eyes."

Mark Twain, *Life on the Mississippi*, 1883

THINGS TO DO

- Computational examples to confirm the model, especially for nonlinear systems.
- Relevance to “limiting” cases like bang-bang and sliding-mode control systems.
- Stabilization of the feedback loop when it contains an internal model of the (unstable) exosystem.
- Topological and metric extensions.

REFERENCES

- W.M. Wonham
  Towards an Abstract Internal Model Principle,
  IEEE Trans. on Systems, Man, and Cybernetics
  SMC **6** (11), 1976, pp.735-740.

- W.M. Wonham, Kai Cai
  Supervisory Control of Discrete-Event Systems,
  in monograph series
  Communications and Control Engineering,
  Springer, 2018.

SPECULATION: WHAT IS "CONSCIOUSNESS"?

COULD WE SAY THAT A CREATURE IS "CONSCIOUS" *IF IT HAS AN INTERNAL MODEL OF ITSELF* ?

TO AVOID AN INFINITE RECURSION, SUCH AN INTERNAL MODEL WOULD NEED TO BE AN

ABSTRACTION OF THE SELF

OPEN PROBLEM FOR YOU TO WORK ON!

THANK YOU