



# Positive feedback in climate: stabilization or runaway, illustrated by a simple experiment

Jean-Louis Dufresne, Marion Saint-Lu

## ► To cite this version:

Jean-Louis Dufresne, Marion Saint-Lu. Positive feedback in climate: stabilization or runaway, illustrated by a simple experiment. Bulletin of the American Meteorological Society, 2016, 97 (5), pp.755-765. 10.1175/BAMS-D-14-00022.1 . hal-01384430

**HAL Id: hal-01384430**

**<https://hal.science/hal-01384430>**

Submitted on 26 Oct 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Positive feedback in climate: stabilization or runaway, illustrated by a simple experiment.**

JEAN-LOUIS DUFRESNE \*

*Laboratoire de Météorologie Dynamique (LMD/IPSL), Centre National de la Recherche Scientifique (CNRS), Université Pierre et Marie Curie (UPMC), Ecole Normale Supérieure (ENS), Ecole Polytechnique (EP), Paris, France*

MARION SAINT-LU

*Laboratoire des Sciences du Climat et de l'Environnement (LSCE/IPSL), Centre National de la Recherche Scientifique (CNRS), Commissariat à l'Energie Atomique (CEA), Université de Versailles Saint-Quentin (UVSQ), Gif-sur-Yvette, France*

## ABSTRACT

The response of the various climatic processes to climate change can amplify (positive feedback) or damp (negative feedback) the initial temperature perturbation. An example of a positive feedback is the surface albedo feedback: when the surface temperature rises, part of the ice and snow melts leading to an increase in the solar radiation absorbed by the surface and to an enhanced surface warming. Positive feedbacks can lead to instability. On Venus for example, a positive feedback is thought to have evolved into a runaway greenhouse effect. However, positive feedbacks can exist in stable systems. This paper presents a simple representation of a positive feedback in both a stable and an unstable system. A simple experimental device based on a scale principle is introduced to illustrate the positive feedback and its stabilization or runaway regimes. Stabilization can be achieved whether the amplitude of the positive feedback declines (e.g. "saturation" of the feedback) or remains constant. The device can also be used to illustrate the existence of tipping points, which are threshold values beyond which the amplification due to feedbacks or the stability of the system suddenly changes. The physical equations of the device are established in the framework of the feedback analysis. Key features to understand why a positive feedback does not necessarily lead to a runaway effect are described. The analogy between the different components of the device and those of the climate system is established. Finally, the contribution of individual feedbacks to the total climate response is addressed. **(Capsule Summary)** To illustrate the concepts of positive feedback, stability, instability, tipping point and saturated feedback we conceive a very simple device based on a scale with two water reservoirs.

**1. Introduction**

At the end of the 19<sup>th</sup> century, Joseph Fourier established that the Earth's surface temperature is primarily driven by a fundamental law, the law of conservation of energy (Fourier 1824, 1837; Pierrehumbert 2004). More specifically, he established that the surface temperature adjusts until the energy budget at the surface reaches equilibrium. He also established that any change of the energy budget at the surface has a direct impact on the surface temperature. A few decades later, Svante Arrhenius used the single-layer model framework suggested by Fourier to quantify each term of the energy budget at the surface and to estimate the energy imbalance due to a change in the atmospheric CO<sub>2</sub> concentration (Arrhenius 1896). He also identified that the major physical phenomenon driving the energy budget adjustment was the Stefan-Boltzmann radiation emission law. Arrhenius also analyzed how water

vapor, clouds and snow may vary with the surface temperature and may impact the energy budget. These estimates were very crude and a major improvement was achieved 70 years later when Manabe and Wetherald (1967) estimated the temperature response to an atmospheric CO<sub>2</sub> increase using a single-column radiative-convective model. These studies and many following ones emphasize the importance of changes in water vapor, clouds, snow and ice extent to estimate the surface temperature change.

The atmospheric water vapor amount, for example, increases with temperature following the Clausius-Clapeyron law assuming fixed relative humidity. As water vapor is a powerful greenhouse gas, such an increase contributes to a further warming of the climate. This positive 'feedback' loop amplifies a given perturbation of the surface temperature by about a factor two (Manabe and Wetherald 1967; Dufresne and Bony 2008). This feedback is amplified as the climate warms due to the increase in the tropopause height (Meraner et al. 2013). Depending on the conditions, this feedback amplification can eventually turn into a runaway greenhouse effect (Kasting 1988; Gold-

---

\*Corresponding author address: Jean-Louis Dufresne, LMD/IPSL, UPMC boîte 99, 4 place Jussieu, 75272 Paris, France  
E-mail: Jean-Louis.Dufresne@lmd.jussieu.fr

blatt et al. 2013): as the temperature rises, the moist convective layer becomes optically thicker and absorbs more and more infrared radiation. A tipping point is reached when the layer becomes fully opaque and the tropopause acquires a fixed temperature (Goldblatt et al. 2013). The surface temperature then increases without modifying the outgoing radiative flux until the planet is so warm that it emits near-visible radiations. A runaway greenhouse effect is thought to have happened on Venus, explaining its current atmospheric composition and its high density (Kasting 1988).

Another climate feedback is due to surface albedo: a surface cooling, for example, increases ice and snow cover, which leads to an increased surface albedo and to a further cooling of the surface (Budyko 1969; Sellers 1969). For a very cold Earth, an extra cooling perturbation can trigger a runaway regime and the Earth can eventually be totally covered by snow: this is the so-called snowball Earth hypothesis (Stern et al. 2006; Kopp et al. 2005).

It could be intuitively stated that a positive feedback "never stops" and is necessarily associated with a runaway regime. This actually may happen in some cases defined as unstable, but this is not necessarily the case. The aim of this paper is to illustrate how systems (e.g. the climate) with positive feedbacks can be stable. When a system is stable, the positive feedback simply amplifies a given perturbation without triggering a runaway regime. The system then stabilizes and reaches a new equilibrium. The physical environment of the system in each of its equilibria is similar. This paper also illustrates how systems with positive feedbacks can be or can become unstable. In this case the positive feedback triggers a runaway regime and the system evolves and reaches a very different state. To investigate positive feedbacks in both a stable and an unstable system, a simple mechanical device is used. This device is based on a scale where a positive feedback can be added. When a weight is added to one side of the scale, the scale either slightly tilts and stabilizes or it turns over, depending on the setup. Simple developments of this device further illustrate the existence of tipping points (i.e. thresholds at which the amplitude of the feedback and potentially the stability are abruptly modified).

The experimental device, its stability, and the interpretation of the results are presented in section 2. The analogy between the experiment and the climate system is presented in section 3, first using a very simple model and then using the classical approach used to analyze comprehensive climate models. Some more details on climate feedbacks and on the question of individual contribution of feedbacks to the temperature change are addressed in section 4. Complementary discussions are given in section 5.

## 2. The mechanical device and the experiment

### a. Description of the device

The device is shown in Fig.1 and Fig.2 gives a simplified schematic of it. The simplifications of the schematic aim at facilitating the mathematical expressions to model the device. The device is a cross attached to a fixed pillar through its intersection  $O$ , so that the cross can rotate around this fixed axis. Initially, the cross is vertical with a long vertical bar and a shorter horizontal bar. One cylindrical water container is hung on each side of the horizontal bar, equidistant from  $O$  and separated by the distance  $l_2$  (note that the containers stay vertical in this simplified schematic whereas in reality they rotate with the device). The containers are about half filled with water and have the same mass. They can communicate through a pipe with a tap at their bottom and another pipe at their top, so that the circulation system is closed. A mass  $m_3$  is at the bottom of the vertical bar, at the distance  $l_3$  from  $O$ .

The tilt  $\Delta\theta$  of the cross is driven by the sum of the moments of force, which are given here by the product between the mass weight and the horizontal projection of the pivot-mass vector.

The construction of the device is of no practical difficulty and very common materials may be used. Feeders for small rodents have been used for the reservoirs and the pipe and tap have been found in aquarium stores.

### b. The experiment

Initially, the system is vertically straight as illustrated in Figs.1a and 2a and the tap between the two containers is closed. At equilibrium, the sum of the moments of force is null. A small mass  $m_1$  is added to one extremity of the horizontal bar, and this leads to a tilt of the device by a small angle  $\Delta\theta_i$  (subscript  $i$  stands for initial change) (Figs.1b and 2b). The angle is small because of the counterweight of the mass  $m_3$ , which generates a return moment.

If the tap is now opened, the water follows gravity and flows from the upper reservoir to the lower one (Figs.1c and 2c). As the lower container becomes heavier and the upper one lighter, the initial tilt of the device is amplified. The height between the two reservoirs increases, more water flows down, and the tilt is further amplified. This is a positive feedback.

How can this amplification loop be stopped? When the tilt increases, the return moment of the counterweight  $m_3$  also increases and balances the growth of the "horizontal bar moment" (i.e. the sum of the moments generated by the masses on the horizontal bar) due to the water transfer. The return moment then stabilizes the device with a final tilt  $\Delta\theta$ .

To understand why the device eventually stabilizes and which conditions need to be met for this, the experiment is described using a simple mathematical model (see details

in the supplementary material). When the stabilization is reached, the final tilt  $\Delta\theta$  is the sum of the initial tilt  $\Delta\theta_i$  due to the addition of  $m_1$  and its amplification  $\Delta\theta_f$  due to the feedback (subscript  $f$  stands for feedback). This amplification  $\Delta\theta_f$  depends on the water mass transferred from the upper reservoir to the lower one. The latter depends on the height between the two containers, which depends on the total tilt  $\Delta\theta$ . In other words, the tilt amplification  $\Delta\theta_f$  depends on the final tilt  $\Delta\theta$  and can be written as  $\Delta\theta_f = g\Delta\theta$  where  $g$  is a function of the containers cross section  $S$ , the water density  $\rho$ , and the device parameters presented above (see supplementary material):

$$g = \frac{\rho S l_2^2}{2 l_3 m_3} \quad (1)$$

The final tilt can be written:

$$\Delta\theta = \Delta\theta_i + \Delta\theta_f = \frac{1}{1-g}\Delta\theta_i \quad (2)$$

where  $g$ , defined by Eq. 1, is now the feedback gain as defined in the classical feedback literature. If  $g$  is positive and smaller than 1, the final tilt at equilibrium  $\Delta\theta$  is larger than the initial tilt without feedback  $\Delta\theta_i$ , as it can be tested using the described experiment. Indeed, according to Eq. 1,  $g$  is necessarily positive, and smaller than 1 as long as the counterweight  $m_3$  is heavy enough. Note that the gain  $g$  can also be increased with the distance between the containers and with their cross section, or decreased with the length of the vertical bar. These parameters are chosen from the beginning and stay fixed once the device is built, while the mass  $m_3$  can easily be adjusted by using a bottle and filling it as much as needed.

When the feedback gain  $g$  becomes close to 1, the final tilt  $\Delta\theta$  becomes infinitely greater than  $\Delta\theta_i$  (cf. Eq. 2), meaning that the system is unstable. In the experiment, this happens if the counterweight  $m_3$  is suppressed or if its mass is too small: the device is not able to stabilize and it keeps on tilting until it eventually turns over, reaching a very different equilibrium (Fig.1d). In practical terms, the mathematical expression of  $g$  is only valid for small tilts. When  $g$  is close to 1, the mathematical expression of  $g$  points out that the system is unstable, but the value of  $g$  cannot describe the final state of the device, as the linear equations describing the system are not valid anymore. The instability consists in a change of regime, as illustrated by the experiment: in the unstable case, the positive feedback generates a runaway from the initial state into a different regime, in which the device can reach a very different equilibrium.

### c. Possible extensions

The device can be adapted to illustrate more complex situations.

**Saturation** The positive feedback loop can be stopped by saturation processes. For example if the upper reservoir becomes empty, the positive feedback stops. On the contrary, if there is too much water, the lower reservoir fills up, which also interrupts the positive feedback loop.

**Tipping point** The containers of the device now have a varying section as illustrated in Fig.3. The cylindrical containers are narrower at their center than at their extremities. Eq.1 shows that increasing the section amplifies the feedback gain. If the water level is initially in the middle of both containers and if the added perturbation is strong enough, the tilt can make the water reach the level of changing section in both containers. This would amplify the feedback. In some cases, this amplitude change can be so large that it can make the system unstable. On the contrary, if the cylindrical containers are widened at their center, the feedback is reduced if the water reaches the level of changing section.

## 3. Analogy with climate

### a. Example of the surface albedo feedback

Let us consider an idealized climate system with only one positive feedback, the so-called albedo feedback. The Earth is considered as a simple isothermal sphere that can be described by global variables (e.g. temperature, radiative fluxes) that do not depend on latitude. There is no atmosphere, no ocean and the surface albedo is assumed to change with the temperature in order to mimic the ice or snow cover change. If the device would represent this climate system, the angle  $\theta$  would be the surface temperature,  $T$ , and the horizontal bar moment would be the absorbed solar flux at the surface,  $F$ . The latter can be perturbed by some external process, as represented by the addition of the mass  $m_1$  in the experiment. The first-order response of the surface temperature to the absorbed solar flux follows the black body emission  $T = (F/\sigma)^{1/4}$  (Planck law), where  $\sigma$  is the Stefan-Boltzmann constant. This is the "Planck response". By analogy, the inclination of the device responds to the horizontal bar moment following the balance with the return moment of the counterweight  $m_3$ .

As the surface warms up, part of the ice or snow covered area melts, reducing the surface albedo,  $\alpha$ , increasing the solar flux absorbed by the surface. Absorbed solar radiation and surface temperature variations thus reinforce each other via a positive feedback. A feedback gain  $g_\alpha$  can be computed (see supplementary material for details):

$$\begin{cases} g_\alpha = -\frac{\lambda_\alpha}{\lambda_P} \\ \lambda_\alpha = -I \frac{\partial \alpha}{\partial T} \\ \lambda_P = -4\sigma T^3 \end{cases} \quad (3)$$

where  $I$  is a quarter of the solar constant,  $\frac{\partial \alpha}{\partial T}$  is the sensitivity of the surface albedo  $\alpha$  to the temperature, and  $\lambda_\alpha$  and  $\lambda_p$  are called the albedo response parameter and the Planck response parameter respectively. The albedo decreases with temperature, meaning that  $\lambda_\alpha > 0$  and  $g_\alpha > 0$  (i.e. the feedback is positive). This system is stable if  $g_\alpha < 1$ , i.e. if the additional absorbed solar energy induced by the decreasing albedo ( $\lambda_\alpha$ ) is smaller than the additional energy loss via radiative emission at the surface ( $-\lambda_p$ ). For current climate conditions, the former is about  $0.3 \text{ Wm}^{-2}\text{K}^{-1}$  and the latter about  $3 \text{ Wm}^{-2}\text{K}^{-1}$  (Dufresne and Bony 2008), indicating that the idealized climate system with only surface albedo feedback is clearly stable. The basic phenomenon that stabilizes the climate, in this very simple model but also in the real climate on Earth and on other planets, is the black body emission as given by the Planck law. The analogue of the Planck law for the device is the return moment of the mass  $m_3$ .

*Saturation and tipping point* The examples of saturated feedback and tipping point presented with the experiment illustrate well some climatic processes that can come into play with the albedo feedback. For example, the albedo feedback saturates if – in case of a surface warming – there is no more ice or snow on Earth. A tipping point may exist if – in case of a surface cooling – the ice-sheet extends equatorward beyond a critical latitude (Budyko 1969; Sellers 1969). Indeed, little variations in albedo have stronger impacts in the tropics than in the polar regions. Passing this tipping point can trigger a snowball Earth state.

#### b. Feedback analysis for the climate system

Over the last few decades, the feedback framework has been used to analyze the climate system (e.g. Hansen et al. 1984; Soden and Held 2006; Bony et al. 2006). This feedback analysis focuses on radiative and short-term feedbacks, as it assumes the climate system is at equilibrium. Some key aspects are presented here to link the experiment with the climate system, where many feedback processes are at play. To do so, the response of the whole climate system to an external perturbation is first considered. This response is then decomposed into a no-feedback response and an additional response due to feedbacks. Finally the feedback gain is introduced.

*(i) State equation.* The classical climate feedback analysis is rooted in a simple non-dimensional linear energy budget model, that gives the global mean temperature change  $\Delta T$  at equilibrium in response to a radiative forcing  $\Delta Q$  (Fig. 4-a):

$$\Delta T = -\frac{\Delta Q}{\lambda} \quad (4)$$

where  $\Delta Q$  ( $\text{Wm}^{-2}$ ) is positive downward (i.e. when it heats the climate system) and  $\lambda$  ( $\text{Wm}^{-2}\text{K}^{-1}$ ) represents

how all the processes of the climate system modify the net flux at the tropopause (positive downward) when the surface temperature is changed.  $\lambda$  is thus negative, as an increase in the surface temperature decreases the net flux at the tropopause. This parameter is called the “climate response parameter”.

*(ii) Introducing a feedback loop.* Based on the current physical understanding of the climate system, the climate response parameter is split into two parts. A classical decomposition consists in isolating the Planck response  $\lambda_p$  as was done in the previous section. This choice will be discussed later. The response parameter  $\lambda_f$  of the remaining processes is then:

$$\lambda_f = \lambda - \lambda_p. \quad (5)$$

This decomposition allows the introduction of a feedback loop. When only the Planck response is considered, the temperature response to a variation  $\Delta F$  of the incoming flux at the tropopause reads:

$$\Delta T = -\frac{\Delta F}{\lambda_p} \quad (6)$$

The flux variation at the tropopause  $\Delta F$  is the sum of the radiative forcing  $\Delta Q$  and a complementary contribution  $\Delta F_f$  due to the feedback processes (included in the response parameter  $\lambda_f$ ):

$$\Delta F = \Delta Q + \Delta F_f \quad (7)$$

$$\Delta F_f = \lambda_f \Delta T \quad (8)$$

Equations 5 to 8 are equivalent to Eq. 4. They describe the whole system after a feedback loop has been introduced as represented on Fig. 4-b.

*(iii) Introducing the feedback gain.* Without any feedback (i.e. when only the Planck response is considered) the temperature change in response to the radiative forcing  $\Delta Q$  is:

$$\Delta T_p = -\frac{\Delta Q}{\lambda_p} \quad (9)$$

The temperature increase  $\Delta T$  is the sum of the Planck response  $\Delta T_p$  and the additional response due to feedbacks  $\Delta T_f$  (Fig. 4-c). The previous equations (5 to 8) can be rewritten to introduce the feedback gain  $g$  used in classical feedback analysis:

$$\Delta T = \Delta T_f + \Delta T_p \quad (10)$$

$$\Delta T_f = g \Delta T \quad (11)$$

$$g = -\frac{\lambda_f}{\lambda_p} \quad (12)$$

The description of the system using this set of equations is represented on Fig. 4-c and the classical expression of the temperature response using the feedback gain reads:

$$\Delta T = \frac{1}{1-g} \Delta T_p \quad (13)$$

The analog of this equation in the experiment is Eq. 2. The initial tilt  $\Delta\theta_i$  is analog to the Planck response, as it corresponds to the initial response when only the counterweight  $m_3$  (analog to the Planck law) acts. The gain  $g$  quantifies how all the feedbacks (i.e. all the processes except those included in the Planck response parameter) reinforce or oppose the Planck response  $\Delta T_P$ .

(iv) *Non linearity of the feedback response.* When the response parameter due to feedbacks is doubled (i.e.  $\lambda'_f = 2 * \lambda_f$ ) the feedback gain is also doubled (Eq. 12). But this is not the case for the total response (Eq. 13) nor for the additional response due the feedbacks (Eq. 11). This can be verified with the experimental device by adding a new pair of identical containers that can communicate through an independent pipe (see supplementary material). When the feedback gain  $g$  is doubled, the response due to a positive feedback (i.e. the increase of the tilt) is multiplied by more than a factor of two. This is a direct consequence of Eq. 2. The amplitude of the additional response due to the feedback depends on the value of  $g$  and is larger as  $g$  becomes closer to one.

#### 4. Climate feedbacks

An overview of how results from climate models are usually analyzed is provided here to understand and quantify the main physical phenomena driving the amplitude of the surface temperature change in response to an external forcing.

##### a. Forcings

The radiative forcing  $\Delta Q$  ( $\text{Wm}^{-2}$ ), introduced in section 3, is defined as the perturbation of the net radiative flux at the tropopause before any surface temperature takes place. The change in the net radiative flux at the tropopause is close to the one at the top of the atmosphere (Hansen et al. 1984; Stuber et al. 2001). This forcing was originally computed assuming no changes in the troposphere and could therefore be computed using radiation-only models. The adjustment of the troposphere to the forcing while the surface temperature remains constant has been recently included in the computation (Gregory and Webb 2008; Andrews and Forster 2008; Boucher et al. 2013; Vial et al. 2013), modifying the exact value of the forcing but not the global climate response. As an example, the current estimate of the radiative forcing for a doubling of the  $\text{CO}_2$  concentration is about  $3.7 \text{ Wm}^{-2}$  (Myhre et al. 2013; Dufresne and Bony 2008).

##### b. Response parameters and climate feedbacks

In classical climate feedback analyses, the Planck response parameter  $\lambda_P$  is chosen as the no-feedback response. In the previous sections where atmosphere and ocean have not been considered, the Planck response was

the first-order response of the outgoing radiative flux to the surface temperature. The Planck response was thus defined unambiguously as the no-feedback response. The choice of this no-feedback response is extended to the real climate and the Planck response is usually defined as the change in the net radiative flux at the tropopause, assuming that the temperature change is vertically uniform from the surface to the tropopause. This choice is arbitrary. It determines the expression and magnitude of the feedbacks, without changing the response of the whole system to a given forcing. Many other choices of the no-feedback parameter can be made. For instance Held and Shell (2012) suggest using a new reference response, where the humidity change required to maintain a fixed relative humidity is considered in addition to the classical troposphere and surface temperature change. The feedback gain, and more generally the analysis of any feedback loop, is not an intrinsic property of the system but the result of how the response of the system has been decomposed into a no-feedback response (or open loop) and an additional response due to feedbacks (i.e. due to closing the loop), and this choice is arbitrary (Ingram 2010; Held and Shell 2012; Lahellec and Dufresne 2013). However, the system will always become unstable when  $g$  becomes close to 1, assuming that the initial perturbation with no feedback is stable.

In classical climate feedback analysis:

$$\begin{cases} \lambda = \lambda_P + \lambda_f \\ \lambda_f = \lambda_w + \lambda_L + \lambda_c + \lambda_\alpha \end{cases} \quad (14)$$

which are respectively the water vapor ( $w$ ), the temperature lapse rate ( $L$ ), the cloud ( $c$ ) and the surface albedo ( $\alpha$ ) response parameters.

These feedbacks have been analyzed in many studies (e.g. Hansen et al. 1984; Soden and Held 2006; Bony et al. 2006; Bates 2007; Soden et al. 2008; Boucher et al. 2013) and are briefly presented here:

- **Water vapor feedback ( $w$ ) and lapse rate feedback ( $L$ ):** The water vapor amount in the atmosphere increases with temperature, leading to an increased absorption of longwave (and shortwave to a lesser extent) radiation and to a decrease of the outgoing flux at the tropopause. The water vapor feedback is thus positive ( $\lambda_w > 0$ ). The temperature in the upper troposphere increases more than the surface temperature in the tropical regions and over most of the globe, which increases the radiation emitted by the upper troposphere to space and contributes to cool the climate. The lapse rate feedback is negative ( $\lambda_L < 0$ ). These two feedbacks are closely linked, particularly in the upper troposphere, owing to the dependence of moist adiabats on temperature in the tropics. They are strongly anti-correlated and it is relevant to consider their sum (Soden and Held 2006). Held and

Shell (2012) have suggested computing the temperature and the temperature lapse rate feedbacks together at fixed relative humidity, with an additional separate relative humidity feedback term

- Surface albedo feedback ( $\alpha$ ): As detailed in section 3-a, the surface albedo feedback is positive ( $\lambda_\alpha > 0$ ).
- Cloud feedback ( $c$ ): A change of cloud fraction has an effect on radiation, which is generally opposite in shortwave and longwave domains. On the one hand, a smaller cloud fraction increases the incoming solar radiation and thus warms the surface. On the other hand, it increases the outgoing longwave radiation and thus cools the surface. On average, models simulate, with a large spread, a decrease of the cloud radiative effect and an increase of incoming radiative flux due to the cloud response ( $\lambda_c > 0$ ) when the surface temperature increases (Soden and Held 2006; Boucher et al. 2013; Zelinka et al. 2013; Vial et al. 2013)

#### c. Contribution of individual feedbacks to the climate response

The definition of the individual contribution of each feedback to the total response ( $\lambda$ ) of the climate to a given forcing ( $\Delta Q$ ) is challenging. The individual contribution of a feedback is not simply given by the difference between the responses with and without this feedback. Indeed, if this addition is performed it is found the sum does not equal the total response. The individual response of a feedback  $x$  has been shown (Dufresne and Bony 2008; Lu and Cai 2009; Lahellec and Dufresne 2013) to be usefully defined as:

$$\Delta T_x = \frac{g_x}{1 - g} \Delta T_P \quad (15)$$

$$g_x = -\frac{\lambda_x}{\lambda_P} \quad (16)$$

Given this definition, the response of all feedbacks together is the sum of the individual responses of each feedback ( $\Delta T_f = \sum_x \Delta T_x$ ) and the total feedback gain is the sum of the individual feedback gains ( $g = \sum_x g_x$ ).

Figure 5a illustrates how the surface warming in response to a doubling of the atmospheric CO<sub>2</sub> concentration can be decomposed into a Planck response and additional responses due to the three following feedbacks: water vapor plus lapse rate, surface albedo and clouds. The results are those obtained with 12 CMIP3 models and have been published by Dufresne and Bony (2008). The Planck response is about 1.2°C and the total response is 3.1°C. Climate feedbacks account for two-thirds of the total surface temperature variation. According to model averages, the water vapor plus lapse rate feedback increase

the temperature by 0.9°C. The cloud feedback increases it by 0.7°C. The surface albedo feedback has the smallest contribution with 0.3°C.

The ratio between the response of feedback  $x$  and the total response does not depend on the other feedbacks. It only depends on the no-feedback response parameter as shown by combining equations 13 and 15:

$$\frac{\Delta T_x}{\Delta T} = g_x. \quad (17)$$

For the water vapor plus lapse rate feedback, this ratio is 0.3 (Fig. 5b). This feedback contributes to 30% of the global mean temperature response regardless of the other feedbacks. The clouds contribute to 22% and the surface albedo to 8% of the total climate response. Note that for scenarios where the amplitude of the forcing smoothly increases in time, the ocean heat uptake can be included, with some assumptions, in the same equilibrium feedback framework as the one presented here (Dufresne and Bony 2008).

## 5. Summary and discussion

An experimental device based on very simple mechanical phenomena has been presented. It illustrated the fact that a system can be stable even if it includes a positive feedback. When describing the processes leading to a positive feedback, the other processes, in particular those ensuring the stability of the whole system, can sometimes be omitted giving the false impression that the positive feedback will never stop and will necessary leads to a runaway regime. In the experimental device, the stability comes from the moment of force of the counterweight. Suppressing it (say by reducing or removing the counterweight) makes the system unstable. In the climate system, the stability comes from the increase of radiative emission with temperature (Planck law). The climate system is currently stable as long as the amplitude of the forcing is not too large.

The classical feedback analysis framework is valid only when the response of the system is linear (i.e. proportional to the forcing, see Eq. 4). The device may also be used to illustrate some non linearities. When the amount of water in the two containers is low, the positive feedback can saturate and become non-existent for larger tilt amplitudes. On Earth, the magnitude of the snow albedo feedback decreases in a warmer climate as the snow cover decreases. When the containers are narrower at their center than at their extremities, the magnitude of the positive feedback may increase when the amplitude of the perturbation increases. This possibly leads to instability. The snow ball Earth and the runaway greenhouse effect on Venus are example of such possible instabilities.

A feedback is not an intrinsic property of a system but rather the result of how the system has been decomposed

into a no-feedback response (or open loop response) and a response with feedbacks (or closed-loop response). This decomposition is driven by our understanding of the system and by the issues to be addressed. For the climate system, the response to a forcing is commonly split into 5 response parameters, each of them being associated to variables that directly influence the radiative flux at the top of the atmosphere. These response parameters can be used to estimate the no-feedback response, the feedback gain and the individual contribution of each feedback to the total response. The Planck law is often used to define the no-feedback response. The individual contribution of feedbacks depends on all feedbacks but the ratio between the individual contribution and the total response has been shown not to depend on the other feedbacks, even though it depends on the no-feedback response parameter. This is therefore an interesting and useful way of quantifying the individual contribution of the feedbacks. It is noteworthy that this individual contribution of the feedbacks cannot be simply defined as the difference between the responses with and without the feedback.

Climate is a complex system and understanding how it changes in response to some forcing is a major challenge. Part of the difficulty originates from the strong interactions between the many processes that exist in the climate system. The feedback analysis is a useful framework as it allows to quantify how the different parts of the climate system contribute to its global response. We hope the experiment proposed in this article help understanding the feedback analysis framework, the interplay of the different parts of coupled systems and the fact that positive feedbacks do not necessarily lead to runaway effects.

**Acknowledgments.** This work was partly supported by the French “Agence Nationale de la Recherche” Clima-Conf project (grant #ANR-10-CEPL-0003). MSL was funded by the “Cimate-KIC”. We thank the anonymous reviewers for their many valuable suggestions and Audine Laurian for editing the English.

## References

- Andrews, T., and P. M. Forster, 2008: CO<sub>2</sub> forcing induces semi-direct effects with consequences for climate feedback interpretations. *Geophys. Res. Lett.*, **35** (4), doi:10.1029/2007GL032273.
- Arrhenius, S., 1896: On the influence of carbonic acid in the air upon the temperature of the ground. *The London Edinburgh and Dublin Philosophical Magazine and Journal of Science*, **41** (251), 237–276.
- Bates, J. R., 2007: Some considerations of the concept of climate feedback. *Q. J. R. Meteorol. Soc.*, **133** (624), 545–560, doi:10.1002/qj.62.
- Bony, S., and Coauthors, 2006: How well do we understand and evaluate climate change feedback processes? *J. Clim.*, **19** (15), 3445–3482, doi: 10.1175/JCLI3819.1.
- Boucher, O., and Coauthors, 2013: Clouds and aerosols. *Climate Change 2013: The Physical Scientific Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, chap. 7, 571–657, doi:10.1017/CBO9781107415324.016.
- Budyko, M. I., 1969: The effect of solar radiation variations on the climate of the Earth. *Tellus*, **XXI** (5), 611–619.
- Dufresne, J.-L., and S. Bony, 2008: An assessment of the primary sources of spread of global warming estimates from coupled Atmosphere-Ocean models. *J. Clim.*, **21** (19), 5135–5144, doi: 10.1175/2008JCLI2239.1.
- Fourier, J., 1837: General remarks on the temperature of the terrestrial globe and the planetary spaces. *American Journal of Science*, **32** (1), 1–20, trad. by E. Burgess.
- Fourier, J.-B. J., 1824: Remarques générales sur les températures du globe terrestre et des espaces planétaires. *Annales de Chimie et de Physique*, **t. XXVII**, 136–167.
- Goldblatt, C., T. D. Robinson, K. J. Zahnle, and D. Crisp, 2013: Low simulated radiation limit for runaway greenhouse climates. *Nature Geoscience*, **6**, 661–667, doi:10.1038/ngeo1892.
- Gregory, J., and M. Webb, 2008: Tropospheric adjustment induces a cloud component in CO<sub>2</sub> forcing. *J. Clim.*, **21** (1), 58–71, doi:10.1175/2007JCLI1834.1.
- Hansen, J., A. Lacis, D. Rind, G. Russell, P. Stone, I. Fung, R. Ruedy, and J. Lerner, 1984: Climate sensitivity: Analysis of feedback mechanisms. *Climate processes and climate sensitivity*, 130–163.
- Held, I. M., and K. M. Shell, 2012: Using relative humidity as a state variable in climate feedback analysis. *J. Clim.*, **25** (8), 2578–2582, doi:10.1175/JCLI-D-11-00721.1.
- Ingram, W., 2010: A very simple model for the water vapour feedback on climate change. *Q. J. R. Meteorol. Soc.*, **136** (646), 30–40, doi: 10.1002/qj.546.
- Kasting, J. F., 1988: Runaway and moist greenhouse atmospheres and the evolution of Earth and Venus. *Icarus*, **74** (3), 472 – 494, doi: 10.1016/0019-1035(88)90116-9.
- Kopp, R. E., J. L. Kirschvink, I. A. Hilburn, and C. Z. Nash, 2005: The paleoproterozoic snowball Earth: A climate disaster triggered by the evolution of oxygenic photosynthesis. *Proc. Natl. Acad. Sci. USA*, **102** (32), 11 131–11 136, doi:10.1073/pnas.0504878102.
- Lahellec, A., and J.-L. Dufresne, 2013: A formal analysis of the feedback concept in climate models, part I: exclusive and inclusive feedback analyses. *J. Atmos. Sci.*, **70** (12), 3940–3958, doi: 10.1175/JAS-D-12-0218.1.
- Lu, J., and M. Cai, 2009: A new framework for isolating individual feedback processes in coupled general circulation climate models. Part I: formulation. *Clim. Dyn.*, **32** (6), 873–885, doi:10.1007/s00382-008-0425-3.
- Manabe, S., and R. T. Wetherald, 1967: Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *J. Atmos. Sci.*, **24** (3), 241–259.



Meraner, K., T. Mauritsen, and A. Voigt, 2013: Robust increase in equilibrium climate sensitivity under global warming. *Geophys. Res. Lett.*, **40** (22), 5944–5948, doi:10.1002/2013GL058118.

Myhre, G., and Coauthors, 2013: Anthropogenic and natural radiative forcing. *Climate Change 2013: The Physical Scientific Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, chap. 10, 659–740.

Pierrehumbert, R. T., 2004: Greenhouse effect: Fourier's concept of planetary energy balance is still relevant today. *Nature*, **432**, 677.

Sellers, W. D., 1969: A global climate model based on the energy balance of the Earth-atmosphere system. *J. Appl. Met.*, **8**, 392–400.

Soden, B. J., and I. M. Held, 2006: An assessment of climate feedbacks in coupled ocean atmosphere models. *J. Clim.*, **19** (14), 3354–3360, doi:10.1175/JCLI3799.1.

Soden, B. J., I. M. Held, R. Coleman, K. M. Shell, J. T. Kiehl, and C. A. Shields, 2008: Quantifying climate feedbacks using radiative kernels. *J. Clim.*, **21** (14), 3504–3520, doi:10.1175/2007JCLI2110.1.

Stern, R., D. Avigad, N. Miller, and M. Beyth, 2006: Evidence for the snowball earth hypothesis in the arabian-nubian shield and the east african orogen. *Journal of African Earth Sciences*, **44** (1), 1 – 20, doi:10.1016/j.jafrearsci.2005.10.003.

Stuber, N., R. Sausen, and M. Ponater, 2001: Stratosphere adjusted radiative forcing calculations in a comprehensive climate model. *Theor. Appl. Climatol.*, **68** (3–4), 125–135, doi:10.1007/s007040170041.

Vial, J., J.-L. Dufresne, and S. Bony, 2013: On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Clim. Dyn.*, **41** (11–12), 3339–3362, doi:10.1007/s00382-013-1725-9.

Zelinka, M. D., S. A. Klein, K. E. Taylor, T. Andrews, M. J. Webb, J. M. Gregory, and P. M. Forster, 2013: Contributions of different cloud types to feedbacks and rapid adjustments in CMIP5. *J. Clim.*, **26** (14), 5007–5027, doi:10.1175/JCLI-D-12-00555.1.

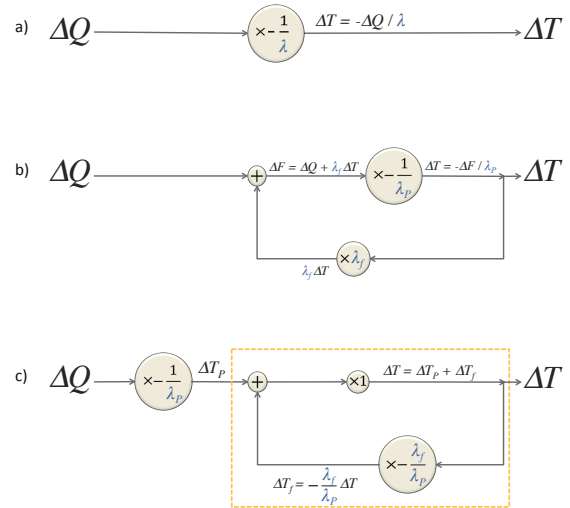


FIG. 4. Three steps in understanding the feedback analysis formalization: (a) considering the change of  $T$  in direct response to the perturbation of  $F$ , (b) introducing a feedback loop, and (c) introducing a feedback gain. Blue fonts indicate the response parameters.

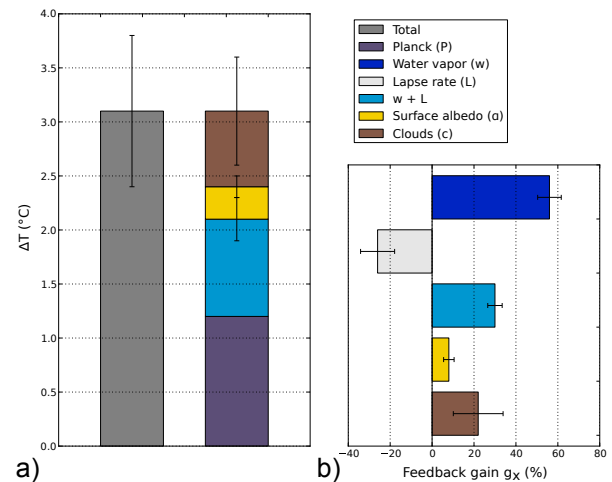


FIG. 5. For a  $\text{CO}_2$  doubling, (a) equilibrium temperature change ( $\Delta T$ ) and contributions to this temperature change due to the Planck response, combined water vapor and lapse-rate (WV + LR) feedback, surface albedo feedback, and cloud feedback; (b) relative contribution of each feedback to the equilibrium temperature change  $\Delta T$  reported in (a). The multi-model mean and the  $\pm 1$  standard deviation for 12 CMIP3 models are shown, using results of Dufresne and Bony (2008)

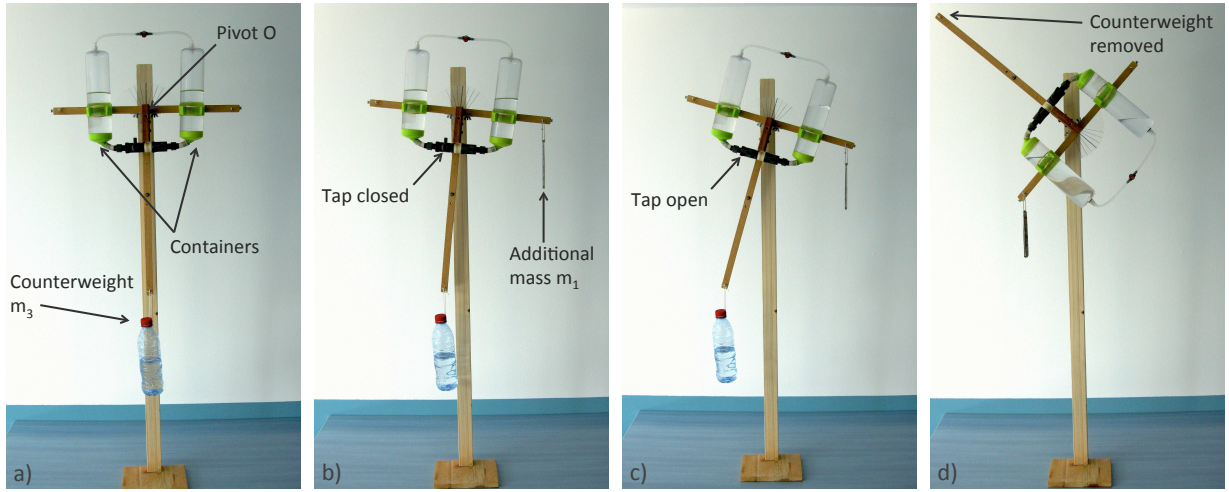


FIG. 1. The experimental device: (a) initial state, (b) perturbed state with closed tap, (c) perturbed state with opened tap, (d) perturbed state with opened tap and without counterweight.

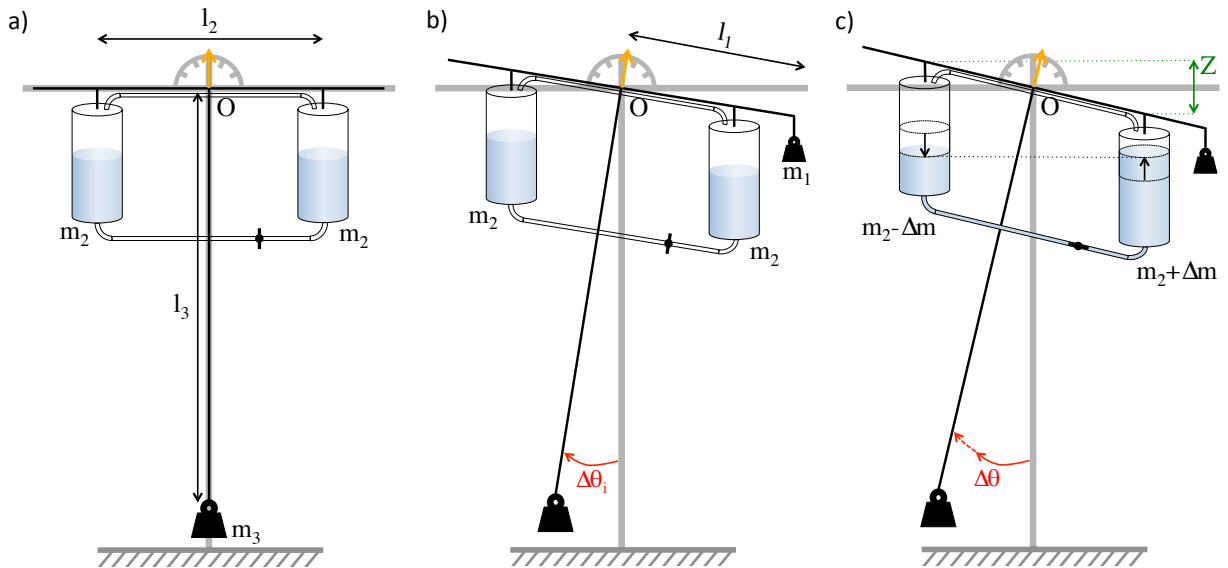


FIG. 2. Schematic of the experimental device and its three states in the stable linear case: (a) initial state, (b) perturbed state with closed tap describing the direct response to an external perturbation, and (c) perturbed state with opened tap describing the response with feedbacks. Unlike the actual device (Fig.1), the containers remain vertical to facilitate the mathematical formulation.

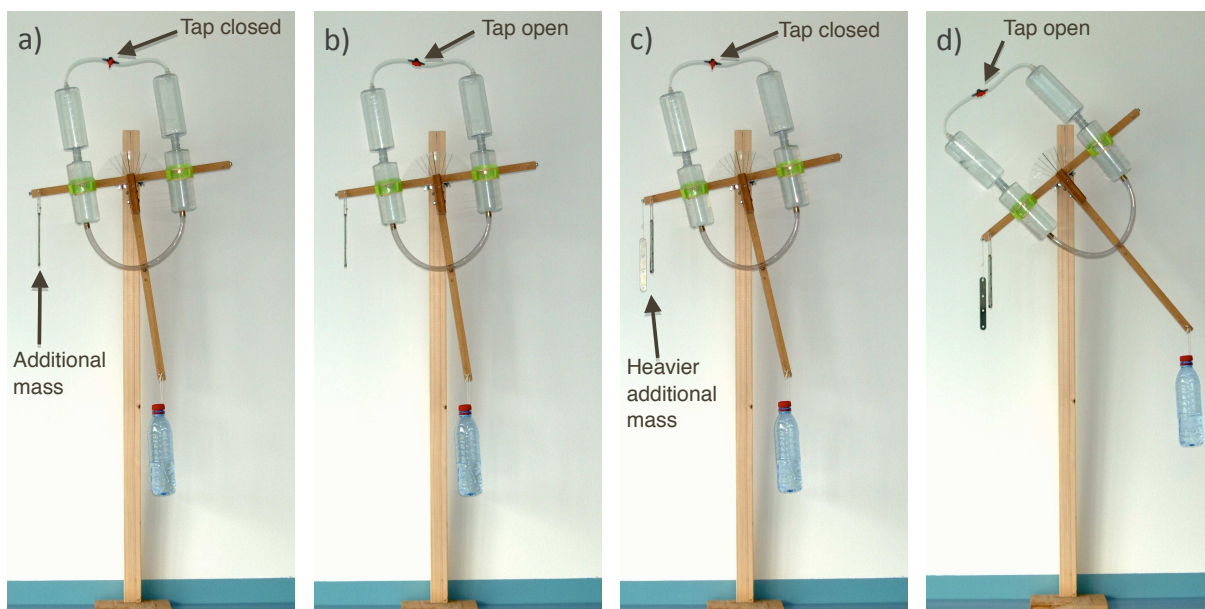


FIG. 3. Modified experimental device shown in Fig.1 where the two containers are narrower at their center than at their extremities, to illustrate a tipping point. The additional mass is larger in (c,d) than in (a,b). The tap are closed in (a, c) and open in (b, d). As long as the additional mass is not too heavy, the water remains in the narrow part of the containers and the feedback gain is small, as illustrated by the small tilt difference between (a) and (b). When the additional mass is large enough, the water reaches the wide part of the containers and the feedback gain becomes large as illustrated by the large tilt difference between (c) and (d).

# Positive feedback in climate: stabilization or runaway, illustrated by a simple experiment.

Jean-Louis Dufresne, Marion Saint-Lu

## Supplementary material

### APPENDIX A

#### Mathematical model of the experiment

A simple model of the device is now presented to better understand and describe the behavior of the experiment.

Each cylindrical water container is initially of mass  $m_2$ . The "horizontal bar moment" is  $M^H$ , and the "return moment" is  $M^R$ . Other definitions are given in the text.

At equilibrium, the sum of the moments of force is null:  $M^H + M^R = 0$ . Initially, as the system is vertically straight, the moments are nulls. Given the definition of the moment of force, if the device is tilted by  $\Delta\theta$  compared to this initial state, the corresponding change in the vertical bar moment is  $\Delta M^R = -g_{grav} l_3 m_3 \sin(\Delta\theta)$ , where  $g_{grav}$  is the gravitational acceleration (absolute value). Considering that  $\Delta\theta$  is relatively small, this simplifies to  $\Delta M^R \approx -g_{grav} l_3 m_3 \Delta\theta$ . Therefore, at equilibrium, the tilt change  $\Delta\theta$  in response to any change  $\Delta M^H$  of the horizontal bar moment reads:

$$\Delta\theta = \frac{1}{g_{grav} l_3 m_3} \Delta M^H \quad (\text{A1})$$

#### a. No-feedback response to an external perturbation

Adding a mass  $m_1$  to one extremity of the horizontal bar makes the device tilt by  $\Delta\theta_i$  ( $i$  for initial) (Fig. 4b) and adds the moments  $\Delta M_i^H = g_{grav} l_1 m_1 \cos(\Delta\theta_i)$ , which simplifies to  $\Delta M_i^H = g_{grav} l_1 m_1$  when considering  $\Delta\theta$  small:

$$\Delta\theta_i = \frac{1}{g_{grav} l_3 m_3} \Delta M_i^H \quad (\text{A2})$$

#### 1) RESPONSE WITH FEEDBACK

The next step consists in opening the tap between the two containers, allowing the positive feedback process between tilt and moment (Fig. 4c). The total variation  $\Delta M^H$  of the horizontal bar moment is given by:

$$\Delta M^H = \Delta M_i^H + \Delta M_f^H \quad (\text{A3})$$

where  $\Delta M_i^H$  is the initial perturbation and  $\Delta M_f^H = g_{grav} l_2 \Delta m$  (subscript  $f$  for feedback) is the amplification due the mass  $\Delta m$  of water which has passed from the upper container to the lower one (with  $\Delta\theta$  small so that  $\cos(\Delta\theta) \approx 1$ ). This mass variation is given by the difference  $Z = l_2 \sin(\Delta\theta) \approx l_2 \Delta\theta$  between the height of the two containers:

$$\Delta m = \rho S \frac{Z}{2} \approx \rho S \frac{l_2 \Delta\theta}{2}, \quad (\text{A4})$$

with  $\rho$  the water density,  $S$  the cross section of the containers. It follows:

$$\Delta M^H = \Delta M_i^H + g_{grav} \rho S \frac{l_2^2}{2} \Delta\theta \quad (\text{A5})$$

The variable  $M^H$  is thus a function of  $\theta$  (Eq. A5) and  $\theta$  is a function of  $M^H$  (Eq. A1). Combining Eqs. A1, A2 and A5, it follows:

$$\Delta\theta = \frac{1}{1 - g} \Delta\theta_i = k \Delta\theta_i \quad (\text{A6})$$

with

$$g = \frac{\rho S l_2^2}{2 l_3 m_3} \quad (\text{A7})$$

and

$$k = \frac{1}{1 - g} \quad (\text{A8})$$

where  $g$  is the feedback gain and  $k$  is the feedback amplification factor. To further understand the meaning of the gain  $g$ , it is worth noting that Eq.A6 can also be written  $\Delta\theta = \Delta\theta_i + g\Delta\theta$ , with  $\Delta\theta_i$  the no-feedback response and  $g\Delta\theta$  the additional response due to feedbacks.

## APPENDIX B

### Mathematical model of the albedo feedback

The solar flux  $F$  absorbed by the surface is

$$F = I(1 - \alpha), \quad (\text{B1})$$

where  $I$  is a quarter of the solar constant and  $\alpha$  the albedo. At equilibrium, the flux absorbed by the surface is equal to that emitted by the surface and the surface temperature is given by  $T = (F/\sigma)^{1/4}$  assuming an emissivity of 1, where  $\sigma$  is the Stefan-Boltzmann constant. For small variations, this equation may be linearized and the temperature varies with the absorbed solar flux as follows:

$$\Delta T = -\frac{\Delta F}{\lambda_P} \quad (\text{B2})$$

With

$$\lambda_P = -4\sigma T^3 \quad (\text{B3})$$

As a comparison with the experiment,  $T$  is the analog of the tilt  $\theta$  of the device,  $F$  is the analog of the moment of force  $M^H$ , and Eq. B2 is the analog of Eq. A1.

#### a. No-feedback response to an external perturbation

The absorbed solar flux is assumed to be perturbed by  $\Delta Q$  (positive or negative) by some external process. In response to this perturbation, the surface temperature directly increases or decreases by  $\Delta T_i$  according to Eq. B2:

$$\Delta T_i = -\frac{\Delta Q}{\lambda_P} \quad (\text{B4})$$

This perturbation corresponds to adding the mass  $m_1$  and keeping the tap closed in the experiment and Eq. B4 is an analog to Eq.A2.

#### b. Response with feedback

The resulting variation  $\Delta F$  of the absorbed solar flux can be decomposed into a first part due the external perturbation  $\Delta Q$  and a second part due to the albedo feedback amplification  $\Delta F_\alpha$ :

$$\Delta F = \Delta Q + \Delta F_\alpha \quad (\text{B5})$$

$\Delta F_\alpha$  depends on the surface albedo variation (Eq. B1) and may be expressed as a function of the surface temperature  $\Delta T$  as:

$$\Delta F_\alpha = \lambda_\alpha \Delta T \quad (\text{B6})$$

with

$$\lambda_\alpha = -I \frac{\partial \alpha}{\partial T} \quad (\text{B7})$$

where  $\lambda_\alpha > 0$  as the albedo  $\alpha$  decreases when the surface temperature increases. Combining Eq.B2 and Eqs.B4 to B6, it follows:

$$\Delta T = \frac{1}{1 - g_\alpha} \Delta T_i \quad (\text{B8})$$

with

$$g_{\alpha} = -\frac{\lambda_{\alpha}}{\lambda_p} \quad (\text{B9})$$

The feedback is positive as  $\lambda_{\alpha} > 0$  and therefore  $g_{\alpha} > 0$ . Equation B8 is an analog to Eq.A6 and Eq. B9 is an analog to Eq.A7.

## APPENDIX C

### Experiment where the gain is doubled

Experimental device with two identical pairs of containers that communicate through two independent pipes. The two taps can be opened one after the other to observe that the tilt increase due to two feedbacks is more than two times the tilt increase due to one feedback.

# Figures

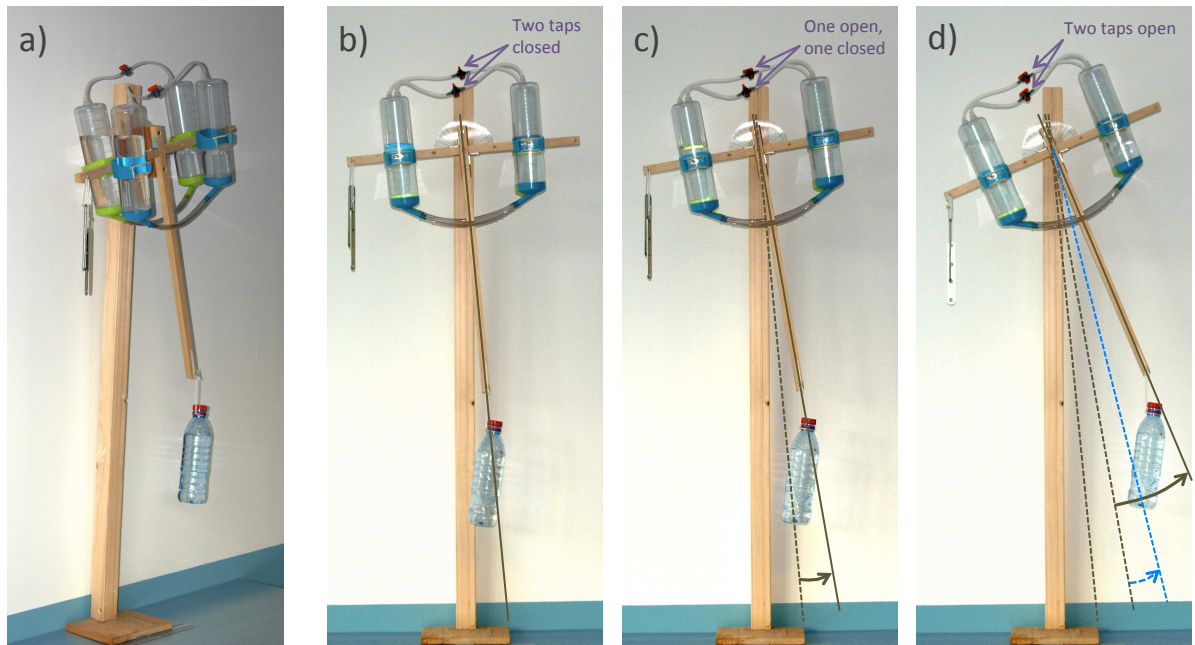


FIG. C1. **Figure C-1:** (a) General view of the device. Perturbed state with (b) the two taps closed, (c) only one tap open and (d) two taps open. On (d), three dashed lines illustrate the tilt with the two taps closed (as in (b)), the tilt when only one tap is open (as in (c)) and the tilt that would be if the additional tilt increase when opening the second tap was identical to the additional tilt increase when opening the first tap (blue line). The difference between the blue line and the actual tilt illustrate that the tilt increase due to two feedbacks is more than two times the tilt increase due to one feedback.